Regularized Multi-Decoder Ensemble for an Error-Aware Scene Representation Network

Tianyu Xiong, Skylar W. Wurster (iD), Hanqi Guo, Tom Peterka, and Han-Wei Shen

Abstract—Feature grid Scene Representation Networks (SRNs) have been applied to scientific data as compact functional surrogates for analysis and visualization. As SRNs are black-box lossy data representations, assessing the prediction quality is critical for scientific visualization applications to ensure that scientists can trust the information being visualized. Currently, existing architectures do not support inference time reconstruction quality assessment, as coordinate-level errors cannot be evaluated in the absence of ground truth data. By employing the uncertain neural network architecture in feature grid SRNs, we obtain prediction variances during inference time to facilitate confidence-aware data reconstruction. Specifically, we propose a parameter-efficient multi-decoder SRN (MDSRN) architecture consisting of a shared feature grid with multiple lightweight multi-layer perceptron decoders. MDSRN can generate a set of plausible predictions for a given input coordinate to compute the mean as the prediction of the multi-decoder ensemble and the variance as a confidence score. The coordinate-level variance can be rendered along with the data to inform the reconstruction quality, or be integrated into uncertainty-aware volume visualization algorithms. To prevent the misalignment between the quantified variance and the prediction quality, we propose a novel variance regularization loss for ensemble learning that promotes the Regularized multi-decoder SRN (RMDSRN) to obtain a more reliable variance that correlates closely to the true model error. We comprehensively evaluate the quality of variance quantification and data reconstruction of Monte Carlo Dropout (MCD), Mean Field Variational Inference (MFVI), Deep Ensemble (DE), and Predicting Variance (PV) in comparison with our proposed MDSRN and RMDSRN applied to state-of-the-art feature grid SRNs across diverse scalar field datasets. We demonstrate that RMDSRN attains the most accurate data reconstruction and competitive variance-error correlation among uncertain SRNs under the same neural network parameter budgets. Furthermore, we present an adaptation of uncertainty-aware volume rendering and shed light on the potential of incorporating uncertain predictions in improving the quality of volume rendering for uncertain SRNs. Through ablation studies on the regularization strength and decoder count, we show that MDSRN and RMDSRN are expected to perform sufficiently well with a default configuration without requiring customized hyperparameter settings for different datasets.

Index Terms—Scene representation network, deep learning, scientific visualization, ensemble learning

---◆---

1 INTRODUCTION

Continuous functional representations of scientific datasets have gained attention as proxies for visualization and analysis thanks to their advantages in compactness, competitive modeling accuracy, and the ability to evaluate values and gradients at random locations without decoding the full volume. A Scene Representation Network (SRN) is a neural functional representation trained with volume data to learn a coordinate-value mapping to attain the aforementioned benefits. For the past several years, the scientific visualization (SciVis) community has improved SRN for volumetric data modeling in its compressiveness [12,16,23,43], computational efficiency [43,48,52], and adaptivity to data [53].

Despite the remarkable advancements in these areas, as a lossy approximation of the data, SRN lacks the ability to depict its prediction quality. Since each value prediction made by SRN exhibits a certain level of error compared to the ground truth, and the most inaccurate predictions are often present in regions with scientific features that involve complex spatial patterns, it is desirable to aid the process with a quantifiable metric that reflects how accurately the SRN reconstructs the data. However, it is non-trivial to obtain such a measurement of the coordinate-wise prediction quality at inference time because evaluating coordinate-wise prediction errors requires the ground truth data, which are often discarded after training. Alternatively, precomputing the error field and writing it to the disk would incur considerable storage costs equivalent to the data size and hence defeats the purpose of using an SRN as a compact surrogate.

In this paper, we propose a method to indicate the prediction quality for SRNs at inference time, which is storage-free and can be evaluated at arbitrary locations. We explore the application of uncertain neural network architectures for computing prediction variances or uncertainties to indicate the confidence level of the model for its predictions. These architectures, including Bayesian and ensemble methods [3, 9, 20], can be used to make multiple plausible predictions for an input such that the variance and mean prediction amounts to a more accurate prediction with the error quantified. In addition to meeting the desired attributes for a proper prediction quality metric, the computed coordinate-level variance along with the mean can be conveniently granted a probabilistic interpretation and integrated into uncertainty-aware visualizations such as probabilistic marching cubes [32] and uncertain volume rendering [33], such that the uncertain predictions can be directly integrated to visualizations as an alternative to visualizing the mean predictions.

To empower SRNs with variance estimation for confidence-aware predictions, we introduce a feature grid SRN architecture with an ensemble of decoders, dubbed multi-decoder SRN (MDSRN), that can seamlessly extend existing feature grid SRNs to produce different plausible predictions for any given input coordinate. Common ensemble methods for variance quantification such as Deep Ensemble (DE) [20] require independent neural networks trained as members, thus resulting in a linear scaling of parameter counts to the number of members. This might not be best suited for volumetric representations in SciVis where the compactness of the model is of great significance. To better adapt ensembling to feature grid SRNs with improved parameter efficiency than DE, we propose a multi-decoder approach, and adapting our MDSRN to an existing feature grid SRN architecture simply requires training multiple MLP decoders along with a shared feature grid encoder. Feature grid SRNs usually have the majority of parameters concentrated on the grid encoder, hence training additional lightweight MLPs only brings a negligible increase in the model size. Comparable approaches that can be applied to SRNs to generate uncertain predictions include the Bayesian neural network [3,9] or directly predicting a variance. However, they can suffer from inferior data reconstruction accuracy, and we show the proposed MDSRN architecture achieves higher reconstruction accuracy with equivalent network size in Sec. 5.1.

Provided that the variance of predictions for each voxel serves as a

• Tianyu Xiong is with The Ohio State University. E-mail: xiong.336@osu.edu
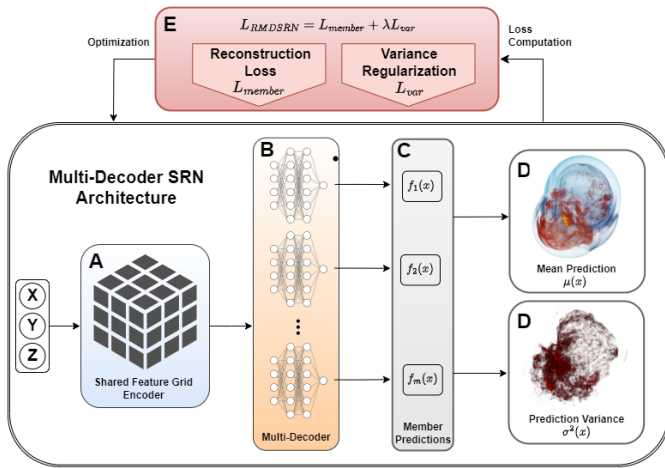
Fig. 1: RMDSRN for volumetric data representation with prediction variance quantification for confidence assessment. A feature grid SRN can be adapted to RMDSRN by training multiple decoders with the reconstruction loss introduced in Sec. 4.1 combing the weighted variance regularization loss detailed in Sec. 4.2 and Sec. 4.3.

metric for assessing the prediction quality, it is imperative for the variance to reflect the accuracy of the mean prediction, where voxels with high prediction errors should also possess high variances. To ensure that the variance from MDSRN is a faithful representation of the prediction quality, we propose a variance regularization loss that synergizes with MDSRN to minimize the distribution discrepancy between the variance and error. The network trained with the regularization, namely the Regularized MDSRN (RMDSRN), exhibits a superior quality of the acquired variance more similar to how error distributes across space. With proper regularization strength, RMDSRN delivers the best data reconstruction and competitive variance quality as an uncertain SRN architecture for error-aware reconstruction and visualization.

To evaluate the efficacy of the proposed MDSRN architecture and the variance-regularized model RMDSRN in providing reliable prediction quality assessment for feature grid SRNs at inference time, we compare with Deep Ensemble (DE), Predicting Variance (PV), and popular Bayesian architectures including Monte Carlo Dropout (MCD) [9] and Mean Field Variational Inference (MFVI) [3]. We present both qualitative visualizations and quantitative results evaluating the spatial similarity of variance and prediction error of all architectures applied to state-of-the-art feature grid SRNs for SciVis [48, 52, 53] across diverse datasets in addition to data reconstruction evaluations. Furthermore, we adapt an uncertainty-aware volume rendering algorithm [33] for uncertain SRNs and study the effect of incorporating uncertain predictions on the rendering quality. To summarize, our contributions are threefold:

- A parameter-efficient architecture for feature grid SRN ensemble, multi-decoder SRN (MDSRN), as a volumetric data representation, that enables post-training prediction quality assessment with coordinate-level prediction variance

- A novel variance regularization loss function used to train a Regularized MDSRN (RMDSRN) so that its prediction variance across space correlates more closely to the actual prediction error

- A comprehensive evaluation of uncertain neural network architectures applied to feature grid SRNs for SciVis. The evaluation spans both data reconstruction and variance estimation with qualitative and quantitative results. The six compared architectures include Monte Carlo Dropout (MCD), Mean Field Variational Inference (MFVI), Deep Ensemble (DE), Predicting Variance (PV), MDSRN, and RMDSRN.

2 RELATED WORK

Our work targets adding error awareness to the predictions and visualizations of scene representation networks (SRNs) with uncertain neural network architectures, and we first review the fast-paced advancements of SRNs for scientific visualization, followed by related methods for uncertainty-aware neural networks.

2.1 Scene Representation Network for Scientific Visualization

Originally proposed as continuous representations of 3D shapes or scenes [24, 28, 39], scene representation networks (SRNs), or equivalently implicit neural representations (INRs), are applied to SciVis as continuous and compact surrogates that go beyond the discretized scientific data formats and support decoding the value at any arbitrary coordinate without full volume reconstruction. A fully implicit approach of SRN that comprises multi-layer perceptrons (MLPs) for scientific data representation was first investigated by Lu et al. [23], and they proposed *neurcomp* that extends the MLP with sinusoidal activation functions known as SIREN [38] with skip connections to achieve a high compression ratio for scalar field data in combination with parameter quantization. SIREN-based SRNs were improved in various aspects following *neurcomp*. Han and Wang [14] showed the versatility of SRN in learning diverse tasks at both data and image levels for spatial-temporal datasets. To increase the computational efficiency of SIREN SRNs, Tang and Wang [43] introduced ECNR that extends the lightweight MLPs organized in a Laplacian pyramid proposed in MINER [34] to work on 4D scientific datasets with deep compression strategies for a high compression rate. Targeting the compression of volume visualization images for spatial-temporal datasets, Gu et al. [12] proposed NeRVI that employs SIREN with skip connections and a CNN upscaling module for faster inference.

As the computational efficiency of MLP-based SRN can be concerning, a hybrid SRN architecture that connects an explicit feature grid with a lightweight MLP, namely feature grid SRN, attracted an interest in SciVis for fast SRNs. Weiss et al. [48] proposed fV-SRN with a composite encoder concatenating features from a dense feature grid and Fourier feature encoding [24, 42] followed by an MLP with SnakeAlt activation, and they further exploited GPU tensor cores for faster inference. Wu et al. [52] achieved interactive training and visualization of hash grid SRN, known as NGP [26], with an optimized training and rendering routine with out-of-core sampling and sample streaming for large-scale datasets. Höhlein et al. [16] explored the effectiveness of multi-grid and multi-decoder SRN in compressing meteorological ensembles. We note the motivation for our multi-decoder ensemble SRN differs from the network by Höhlein et al. as they use the different decoders to learn different volumes in an ensemble simulation dataset, whereas our method applies to a single scalar field for variance quantification as a realization of ensemble neural network. Farokhmanesh et al. [6] modeled bivariate correlations in ensemble datasets with a bipartite hash grid SRN dubbed NDF. For better generalization to unseen inputs, Wu et al. [51] presented HyperINR with a hypernetwork on hash grids trained with knowledge distillation. As scientific data can exhibit certain degrees of sparsity, Wurster et al. [53] proposed APMGSRN to concentrate SRN parameters to regions potentially containing scientific features with high errors through learned transformations for multiple feature grids.

2.2 Uncertain Neural Network Architecture for Prediciton Variance Quantification

The critical necessity of understanding the trustworthiness of neural network predictions spurred a rich set of research on uncertainty-aware neural network architectures [10]. These methods can be applied to neural networks such that an uncertain prediction with variance quantification can be obtained for regression tasks. Ensemble and Bayesian Neural Network (BNN) are two prominent methods for this purpose. Monte Carlo Dropout (MCD) [9] and Mean Field Variational Inference (MFVI) [3] are two practical realizations of BNN. As for the ensemble approach, Deep Ensemble (DE) [20] advocated
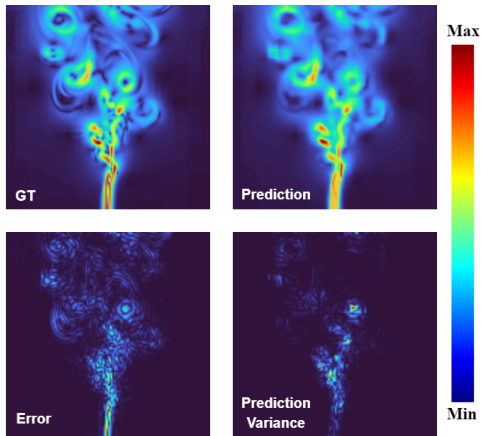
Fig. 2: Illustration of the undesired overconfident variance problem of uncertain SRNs. Despite being the most inaccurate in the bottom thin structure as the error image shows, an uncertain SRN can fail to capture it with the highest variance elsewhere in the domain.

ensembling neural networks for uncertainty quantification with competitive quality compared to Bayesian methods. Brief summaries of these uncertain neural network methods are included in Sec. 3 as they constitute competitive alternatives of variance quantification methods to be evaluated along with ours in Sec. 5.1. Since independent neural networks are trained, DE is parameter intensive, and there is a series of follow-up works focusing on parameter-efficient ensemble for neural networks [21, 44, 45, 49]. Our work shares a similar idea to sub-ensemble [45], where efficiency is achieved by sharing parameters between members. While sub-ensemble is a generic method of parameter-sharing, we provide a specific strategy for the decomposition of the shared and unique parameters targeting feature grid SRNs. In addition, our training routines also differ in ways that we train the shared feature grid jointly with all member decoders, whereas sub-ensemble uses sequential training of members with the shared component frozen after trained with the first member.

In the context of SRNs, several uncertainty-aware SRNs have been studied for different applications. For Computed Tomography (CT) image representation, UncertaINR [46] proposed to use MCD for uncertain SRN. For neural radiance fields (NeRFs), a variance estimation technique used by ActiveNeRF [27] and U-NeRF [4] is to predict the variance as part of the network output in addition to a value prediction, and we refer to this as the Predicting Variance (PV) method as reviewed in Sec. 3.2. S-NeRF [37] proposed to learn independent logistic normal distributions of color and densities with variational inference, and CF-NeRF [36] improved over S-NeRF with a conditional normalizing flow to learn a more flexible distribution of radiance field. Sünderhauf et al. [41] applied DE-based NeRF and formulated a density-aware uncertainty term to augment the RGB uncertainty. Bayes' Rays [11] proposed a Laplacian approximation of the parameter distribution of NeRF relying on a parametric perturbation field. Despite different uncertain NeRFs being proposed, their training methods are often tailored to the NeRF pipeline, posing challenges to applications on other SRN tasks.

3 BACKGROUND

As we propose to equip feature grid SRNs with prediction variance quantification to indicate the quality of the prediction, we briefly review SRNs with feature grid encoders to which our method in Sec. 4 is applied as well as the related variance quantification approaches for neural networks that are compared in Sec. 5.1.

3.1 Feature Grid Scene Representation Networks

Extending the general discussion of SRNs for SciVis in Sec. 2.1, we formally define SRNs and highlight important properties of its feature grid variants. A scene representation network (SRN) is a neural network that predicts scalar or vector values from input coordinates, and its

architecture can be conceptually divided into two components, an encoder E that transforms an input coordinate to a feature and a decoder D that predicts the value given the feature. Formally, let f be an SRN and (x, y) be a coordinate-value mapping, the task of an SRN is to reconstruct y given x: $f(x) = D(E(x)) = y$.

An initial application of SRNs to scientific data was explored by Lu et al. [23]. They proposed *neurcomp* that uses multi-layer perceptrons (MLPs) for both the encoder and the decoder. Despite the high compressiveness achieved by *neurcomp*, it is slow to evaluate as it consists of many wide fully connected layers. To improve the computational efficiency of SRNs, Weiss et al. [48] proposed fV-SRN that substitutes the MLP encoder with a feature grid. A feature grid has learnable features defined on each voxel, and an input coordinate can be encoded with a trilinear-interpolated feature from neighboring voxels. The grid encoder can be evaluated more efficiently than an MLP which requires a chain of matrix multiplications. Furthermore, the decoder can also be made with fewer layers and neurons. As a result, the feature grid SRN can achieve considerable inference speedups. Following this conceptual model of grid encoders with lightweight decoders, more advanced feature grid SRNs are proposed to further improve the efficiency and accuracy for large-scale scientific data, such as Neural Graphics Primitives (NGP) [26, 52] and the Adaptively Placed Multi-Grid SRN (APGMSRN) [53]. In Sec. 4, we introduce our approach that can add to feature grid SRNs the ability of reliable prediction variance quantification with a modification to the given network architecture along with a novel loss function, and our methods integrate well with state-of-the-art models like fV-SRN, NGP, and APMPSRN.

3.2 Prediction Variance Quantification for Neural Networks

In addition to our work that proposes a customized strategy to modify the feature grid SRN architecture for variance quantification, there are alternative methods that can also be applied to achieve similar goals. Specifically, we briefly review Monte Carlo Dropout (MCD) [9], Mean Field Variational Inference (MFVI) [3], Deep Ensemble (DE) [20], and Predicting Variance (PV) [27] before evaluating them in Sec. 5.1.

Similar to our methods, MCD, MFVI, and DE enable the network to produce multiple predictions to an input, from which a variance can be computed. Although dropout is initially proposed to reduce overfitting by randomly turning off neurons according to some probability [40], Gal and Ghahramani [9] proposed to keep dropout at inference time such that different predictions can be acquired with multiple sets of network weights after dropout trails, and this MCD method bears a Bayesian neural network interpretation. MFVI [3] is another Bayesian method, and it requires the network to learn a posterior Gaussian distribution for every weight, such that the network weight distributions can be sampled. Different samples of the network will in turn predict differently. DE [20] is a closely related method to ours as both approaches belong to the ensemble technique for uncertainty quantification. Instead of training one network, DE independently trains an ensemble of networks, such that the variance of the predictions from each member network can be calculated as uncertainty.

PV has a different realization of variance quantification compared to the abovementioned methods in that the network directly predicts the variance, and it was applied to SRNs, specifically neural radiance fields (NeRFs), by Pan et al. [27]. A PV network outputs both a mean and a variance prediction, effectively predicting a Gaussian distribution, and it is optimized with the negative log-likelihood (NLL) loss from the predicted Gaussian and the ground truth.

4 REGULARIZED ENSEMBLE SCENE REPRESENTATION NETWORK

Feature grid scene representation networks (SRNs) have received multitudes of improvements in computational efficiency and accuracy as surrogates for large-scale scientific data, yet there remains the question of where in the domain predictions can be trusted to be accurate such that the volume visualization will be in high quality with respect to the ground truth.

We explore extending feature grid SRNs with a network architecture that produces multiple predictions for every input to acquire prediction

variance as a measurement of prediction quality and trustworthiness, which has been studied in diverse deep learning tasks [5, 13, 17, 25, 35]. Our approach shown in Fig. 1 includes a parameter-efficient architecture for an ensemble of feature grid SRNs detailed in Sec. 4.1 as well as a novel variance regularization loss in Sec. 4.2. We then introduce the regularization strength scheduler that helps a more robust training to the choice of λ in Sec. 4.3.

4.1 Multi-Decoder Ensemble SRN Architecture

Deep Ensemble (DE) [20] provides high-quality variance quantification and predictive performance to neural networks by training an ensemble of member networks to acquire diverse predictions. Despite the attractive performance of DE, an ensemble of independent member networks can be parameter-intensive with a linear scaling factor to the number of members, which can lead to less optimal performance under a constrained parameter budget as in SRN tasks. To address the poor scaling of parameters of DE, we propose a parameter-efficient ensembling strategy for feature grid SRNs, dubbed multi-decoder SRN (MDSRN), that shares parameters between member networks. When constructing an ensemble of feature grid SRNs, all members of an MDSRN model share the feature grid so that none of them learn separate grids. Consequently, the one shared feature grid can use higher resolutions and larger feature sizes that all members can benefit from. Apart from the shared grid encoder, we do not share any layers in the multi-layer perceptron (MLP) decoder and instead use completely separate MLPs for different members. This is to prevent an excessive level of similarity between member networks, which can lead to a spatially homogenous variance that fails to distinguish regions that are challenging versus easy to learn.

We formally introduce the proposed MDSRN architecture that adds variance quantification to feature grid SRNs. As illustrated in Fig. 1, to incorporate our multi-decoder strategy to a feature grid SRN, while the encoder in part (B) requires no modification, an ensemble of decoders in part (C) needs to be trained to output multiple predictions such that a mean and a variance can be computed in part (D) for quality-informed reconstruction and visualization. Following the notations in Sec. 3.1, for an MDSRN containing M member predictors that share the same encoder with different decoders, denoted $f_i(x) = D_i(E(x))$ where $i = 1, 2, .., M$, the mean prediction and the variance are computed as follows:

$$\mu(x) = \frac{\sum_{i=1}^{M} f_i(x)}{M} \qquad \sigma^2(x) = \frac{\sum_{i=1}^{M}(f_i(x) - \mu(x))^2}{M-1} \qquad (1)$$

The design is inspired by the observation that the bottleneck limiting the accuracy of an ensemble of SRNs is usually the capacity of the member networks. For example, for the same number of total parameters, an ensemble of 3 large independent SRNs often has a more accurate mean prediction than an ensemble of 5 smaller independent SRNs. This reveals a key to increasing the accuracy of an ensemble in a fixed compression level is to increase the capacity of member SRNs, and MDSRN achieves this through parameter-sharing. The multi-decoder method can be applied easily to state-of-the-art feature grid SRNs for scientific data including APMGSRN [53], NGP [26, 52], and fV-SRN [48] for post-training prediction confidence evaluation as MDSRN simply requires an ensemble of decoders of their specified architectures, such as MLP with ReLU activation for APMGSRN or SnakeAlt activation [48] for fV-SRN, despite that as a limitation of the ensemble method, training time can scale with the decoder members as shown in Sec. 5.1.

To optimize an MDSRN under the ensemble learning scheme similarly to DE networks, each member is supervised with the ground truth data to ensure they can make sufficiently plausible predictions. Let x and y be a coordinate-value pair and B denote the number of pairs in a training batch, the per-member loss function L_{member} to train MDSRN in part (E) of Fig. 1 is defined as a sum of mean squared errors between each member's prediction and the ground truth:

$$L_{MDSRN} = L_{member} = \frac{1}{B}\sum_{i=1}^{M}\sum_{b=1}^{B}(f_i(x_b) - y_b)^2 \qquad (2)$$

4.2 Variance Regularization Loss Function

Although MDSRN provides improved parameter efficiency for better data reconstruction accuracy than the conventional network ensemble approach at the same compression level, we found the spatial pattern of variance does not always align with that of error, as illustrated in Fig. 2, and this can be problematic when the variance is utilized to evaluate the prediction quality. To mitigate the issue, our idea is to regularize the model with an additional loss function so that the regularized MDSRN (RMDSRN) learns to promote higher similarity between the variance and error. In addition, the adjustments to variance should be ideally based on its own scale to avoid equating two different quantities with potentially disparate value ranges.

We propose to minimize the dissimilarity of the variance and error in the context of probability distributions to achieve scale-invariant matching. By normalizing the variance and error fields such that both quantities integrate to one over the spatial domain, we can obtain valid probability density functions (PDFs) in 3D. Furthermore, since the density of a value at a location indicates the relative magnitude of the value compared with others over the domain, locations with overconfident variance can be naturally identified by those with lower variance densities than the error densities. Consequently, minimizing the difference between the two density functions effectively adjusts the variance in a scale-invariant way.

To acquire proper densities for the variance and error, we can scale each of the values by their total aggregates in the volume. Formally, let X denote all coordinates encompassed by the volume extent and $v(x) = y$ be a function that returns the ground truth value at a location, the variance density function f_{σ^2} and the error density function f_δ are defined as:

$$f_{\sigma^2}(x) = \frac{\sigma^2(x)}{\int_X \sigma^2(x_j)dx_j} \qquad f_\delta(x) = \frac{(\mu(x) - v(x))^2}{\int_X (\mu(x_j) - v(x_j))^2 dx_j} \qquad (3)$$

Although Eq. (3) describes the density functions in the continuous space over the volume for SRNs as continuous representations, only random voxel samples are available in a training batch from an optimization step. To compute the densities during training, we normalize the variances and errors of the sampled points in a batch by their respective sums, which is equivalent to replacing the integrals in Eq. (3) by a summation over voxels in a batch, such that the densities constitute discrete approximations of Eq. (3). The batch-wise densities are fast to compute and can update the model parameters for every optimization step to work with the member reconstruction loss, and results in Tab. 1 show this approach can effectively align the variance of RMDSRN with the error across the volume without increasing the training time significantly compared to MDSRN.

After density functions are defined, we minimize their difference with Kullback–Leibler (KL) divergence as a measure of dissimilarity between probability distributions, and Eq. (4) shows the variance regularization L_{var} used during training. KL divergence especially helps RMDSRN prioritize matching the variance for high-error regions that can greatly affect the accuracy of visualizations, and it is accomplished by weighing the log-space density difference by the error density, hence points with high errors contribute more to the loss. We note that gradients incurred in the computation of error densities are not recorded, because we would like member models to minimize KL divergence solely by steering the variance of their predictions toward similar patterns to error, so gradients should only flow from the loss function to the variance densities.

$$L_{var} = \frac{1}{B}\sum_{j=1}^{B} f_\delta(x_j) log(\frac{f_\delta(x_j)}{f_{\sigma^2}(x_j)}) \qquad (4)$$

By adapting our multi-decoder design and the variance-regularized training as Fig. 1 presents, feature grid SRN can achieve reliable variance quantification. The loss function of the final Regularized MDSRN (RMDSRN) becomes a weighted sum of the member loss and the variance regularization : $L_{RMDSRN} = L_{member} + \lambda L_{var}$.
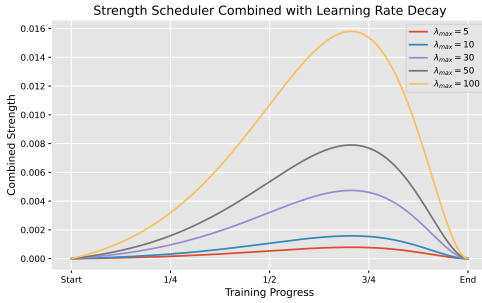
Fig. 3: Visualization of the combined variance regularization strength, which is the learning rate decayed with cosine annealing multiplied by $\lambda(t)$ from Eq. (5), at each step of training with varying λ_{max} for the scheduler.

4.3 Training Regularized MDSRN with Exponential Growth Weight Scheduler

Compound loss functions usually benefit from appropriate weights on each term for the best results, such as in image generation networks like InSituNet [15] and FoVolNet [2]. The quality of a trained RMDSRN model is also dependent on a proper regularization strength λ. We use a λ scheduler to provide more robust training without requiring a hyperparameter search as the detailed ablation study in Sec. 5.3 demonstrates. The RMDSRN trained with the scheduler exhibits improved overall performance than the unregularized MDSRN for a wide λ range as well as the default recommendation.

The scheduler defined in Eq. (5) gradually increases λ from a minimal value to a maximum at an exponential rate during training. The principle behind our λ scheduling is to silence the regularization in early iterations and scale it up as the model approaches convergence in the member loss. The motivation is that since model predictions and hence errors will change after each optimization step, distribution matching becomes a moving target. On the other hand, the true targets that the variance should learn to correlate are the errors in later training iterations, as they do not vary significantly and are more consistent representations of the final RMDSRN error at inference time. With this principle, our scheduler starts λ with a small value ideally close to zero and keeps it small in the early iterations of training to allow a faster progression of the model accuracy. Following exponential growth rates, in later stages of training, the scheduler can quickly raise λ to allow a satisfactory fit to the final error pattern. The weight on training step t is defined as:

$$\lambda(t) = \lambda_{min} + (\lambda_{max} - \lambda_{min})\frac{r^{\frac{t-1}{t_{max}-1}} - 1}{r - 1} \qquad (5)$$

At step 1, the weight is the minimum value λ_{min} and it increases at an exponential rate to λ_{max} at the last training step t_{max}. Here λ_{max} is a hyperparameter to be tuned for the best result of variance regularization, while for λ_{min} a default of zero works generally. The parameter $r \in (1, \infty]$ controls the growth rate of the scheduler. Rates close to 1 result in linear-like growths and the higher the rate, the steeper the growth will be in later iterations and the less weight for early iterations.

As the strength scheduler is applied jointly with learning rate decay, the actual weight on the regularization is $\lambda(t)$ in Eq. (5) multiplied by the current learning rate. We visualize in Fig. 3 the combined strength throughout training with a cosine annealing learning rate scheduler [22] used in Sec. 5.1 with an initial learning rate of 5.0e-3. For all λ_{max} shown, λ_{min} and r are set to 0 and 500. The combined strength grows slowly in early training and progresses quickly to a maximum before 3/4 of training, followed by decreasing values until the end of training. This follows our motivation to scale λ up in later training stages, and the shrinking λ towards the end helps the regularization converge.

In summary, λ_{max} of the scheduler can be tuned for the best results from variance regularization. For λ_{min}, setting it to zero should fit most training scenarios. Regarding the growth rate r, we recommend a default r of 500 so the weight does not grow significantly until around 1/4 of the total training steps as visualized in Fig. 3. A detailed ablation

Table 1: Quantitative data reconstruction and variance evaluation results. For each dataset, evaluated methods include the non-uncertain SRN (as in "SRN") and uncertain models with a specified base SRN architecture. All models in the same dataset have the same size as indicated in the Base SRN column. PSNR evaluates data-level reconstruction accuracy. Pearson correlation (corr) and our modified Jaccard Index with spatial tolerance (JI-ST) assess the spatial similarity between the prediction variance and error. Negative log likelihood (NLL) evaluates the quality of Gaussian distributions parameterized by prediction means and variances in representing the data. The top two statistics among uncertain models in each dataset are bolded with the best one underlined.

Dataset	Base SRN	Model	PSNR (dB) ↑	Corr ↑	JS-ST(1%/5%) ↑	NLL ↓	Training time
Plume 512×128² 32 MB	APMGSRN 0.35 MB	SRN	49.66	N/A	N/A	N/A	404s
		MCD	44.73	0.2928	43.9% / 54.9%	8.28	**430s**
		MFVI	46.00	0.3172	47.0% / 50.1%	-0.59	495s
		DE	46.07	0.3726	40.8% / 52.9%	5.86	804s
		PV	42.53	**0.5415**	**61.9% / 76.9%**	**<u>-5.88</u>**	<u>423s</u>
		MDSRN	**46.66**	0.3521	44.6% / 59.5%	759.08	760s
		RMDSRN	<u>**47.58**</u>	**0.6154**	60.1% / 64.3%	**-4.27**	780s
Nyx 256³ 64 MB	fV-SRN 3.21 MB	SRN	39.39	N/A	N/A	N/A	346s
		MCD	37.09	0.1353	25.7% / 34.1%	6.42	<u>**386s**</u>
		MFVI	37.70	0.0963	20.6% / 30.8%	3.46	494s
		DE	33.98	0.0702	12.5% / 24.9%	-0.85	789s
		PV	37.33	<u>**0.4536**</u>	<u>**50.5% / 54.3%**</u>	<u>**-3.06**</u>	402s
		MDSRN	**39.01**	0.0215	11.3% / 27.9%	418.96	730s
		RMDSRN	<u>**39.04**</u>	0.2134	27.7% / 41.4%	**-1.16**	746s
Supernova 432³ 308 MB	APMGSRN 4.33 MB	SRN	50.36	N/A	N/A	N/A	346s
		MCD	43.52	0.3578	38.9% / 59.4%	-3.72	<u>**386s**</u>
		MFVI	46.83	0.2505	32.8% / 55.9%	-3.98	313s
		DE	46.76	0.4268	47.6% / 60.3%	<u>**-7.48**</u>	789s
		PV	46.01	**0.5364**	**64.6% / 74.0%**	-6.46	**390s**
		MDSRN	**49.31**	0.2089	32.2% / 49.2%	25.66	730s
		RMDSRN	<u>**49.35**</u>	0.5026	59.3% / 68.8%	**-7.05**	746s
Asteroid 1000³ 3815 MB	NGP 40.21 MB	SRN	45.71	N/A	N/A	N/A	188s
		MCD	39.64	0.1242	19.6% / 56.7%	-3.40	**209s**
		MFVI	40.85	0.1609	31.6% / 44.4%	-4.10	417s
		DE	36.19	0.4575	66.2% / 91.2%	<u>**-7.87**</u>	405s
		PV	38.43	<u>**0.6213**</u>	<u>**79.4% / 87.3%**</u>	-7.51	<u>**192s**</u>
		MDSRN	45.86	0.2455	65.4% / 90.8%	-4.57	372s
		RMDSRN	<u>**46.12**</u>	0.4618	73.5% / <u>**91.6%**</u>	-7.59	345s
Isotropic 1024³ 4096 MB	fV-SRN 40.21 MB	SRN	39.57	N/A	N/A	N/A	407s
		MCD	36.44	0.0738	10.2% / 20.9%	7.82	**440s**
		MFVI	34.31	0.0373	5.4% / 12.6%	4.89	529s
		DE	34.99	0.2083	16.4% / 26.9%	-0.33	813s
		PV	36.99	**0.4901**	**44.2% / 55.0%**	<u>**-3.24**</u>	<u>411s</u>
		MDSRN	**39.48**	0.0200	7.6% / 18.1%	1616.72	776s
		RMDSRN	**39.48**	0.2876	29.8% / 41.4%	**-1.68**	790s

study of λ_{max} and the scheduler can be found in Sec. 5.3.

5 EXPERIMENTS

To study the performance of uncertain neural network architectures on feature grid SRNs for scientific data, we evaluate MDSRN, RMDSRN, Deep Ensemble (DE) [20], Monte Carlo Dropout (MCD) [9], Mean Field Variational Inference (MFVI) [3], and Predicting Variance (PV) in reconstruction accuracy and variance quality with evaluation metrics designed for volume visualization along with qualitative results in Sec. 5.1. Related methods are reviewed in Sec. 3.2. We then study an adaptation of uncertainty-aware volume rendering [33] for uncertain SRNs in Sec. 5.2. Finally, we conduct ablation studies of the regularization strength λ on RMDSRN as well as the ensemble member count for both MDSRN and RMDSRN on Sec. 5.3 and Sec. 5.4.

5.1 Uncertain Neural Network Architecture Evalaution

We quantitatively and qualitatively evaluate our methods, MDSRN and RMDSRN, against Deep Ensemble (DE), Monte Carlo Dropout (MCD), Mean Field Variational Inference (MFVI), and Predicting Variance (PV) reviewed in Sec. 3.2. The uncertain architectures are adapted to three state-of-the-art feature grid SRNs for five scalar field datasets with different feature patterns and sparsity. We first detail the experiment settings, and then introduce the evaluation methods and results for the data reconstruction and variance quality tests.

Evaluation datasets and base SRNs. The evaluated datasets include Plume, Nyx, Supernova, Asteroid, and Isotropic. The dimensions of the
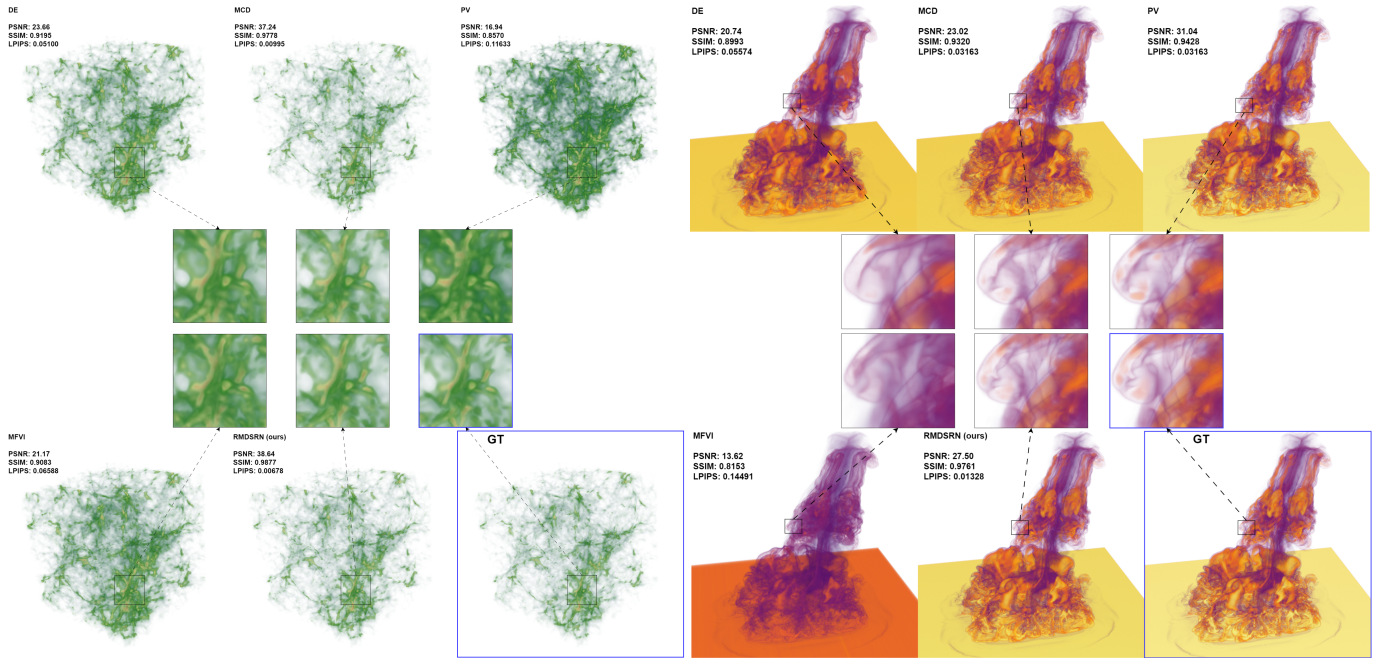
Fig. 4: Volume renderings of reconstructed data for Nyx (left) and Asteroid (Right) as well as the ground truth in blue borders. Both image-level evaluation metrics and the enlarged views show the renderings from RMDSRN maintain the highest fidelity overall and reproduce feature structures the most clearly as evidenced by SSIM and the perceptual loss.

dataset can be found in the first column of Tab. 1 containing quantitative results. We apply uncertain network architectures to state-of-the-art feature grid SRNs for SciVis, including fast Volumetric SRN (fV-SRN) [48], Neural Graphics Primitives (NGP) [26,52], and Adaptively Placed Multi-Grid SRN (APMGSRN) [53]. We pair each dataset with the base SRN whose architecture fits the data feature well, as presented in Tab. 1. For Nyx and Isotropic with dense features over the domain, we use fV-SRN. Plume and Supernova contain features that are uniformly distributed in a subregion, and these features can be properly learned with APMGSRN. As for Asteroid with a higher degree of sparsity, NGP is applied. Please find the exhaustive quantitative results covering uncertain architectures on all base models for every dataset in the supplemental material.

Baseline uncertain neural network architectures. Compared methods are reviewed in Sec. 3.2. For SRNs with MCD and MFVI, the MLP decoder is adapted to their methods and the feature grid encoder is unchanged. For DE, independent feature grid SRNs are trained. To implement PV, a feature grid SRN needs an additional output channel for the decoder to predict the variance, and the NLL loss is applied.

SRN training and hyperparameters. For each pair of datasets and base SRNs, the uncertain SRNs have the same size from 1% to 6% of the data size for a sufficient reconstruction quality. For all uncertain models except for DE, the same encoder configurations are used, whilst the decoder setup is different to achieve an equivalent size since increasing the grid resolution and feature size can easily result in unbalanced parameter counts. MCD and PV each has a 3-layer MLP of 128 neurons, and MFVI has 2 layers of 108 neurons. Both MDSRN and RMDSRN have 5 decoders with 2 layers and 64 neurons. While the 5 members of DE all have a 2-layer decoder with 64 neurons, the encoder capacity is less than other methods for equivalent network sizes.

We implement all models and perform training with PyTorch [29] with an NVIDIA A100 GPU and an AMD EPYC 7643 processor. Training settings are standardized for most models and datasets. We train most models with the Adam optimizer [19] and a learning rate of 5.0e-3 which delivers smooth convergences. For PV, we apply 5.0e-4 as the learning rate following ActiveNeRF [27] as an existing study of PV SRN, and MFVI also uses 5.0e-4 to avoid unstable training. We also apply cosine annealing [22] for learning rate decay with the minimum rate set to 1.0e-7. We randomly select one batch of 2^{17} or 131,072 coordinate-value pairs in every training iteration for 50000 steps.

For uncertain architectures with hyperparameters, 3 trials of training with different settings are performed. For MCD, we test the dropout probability $p \in \{0.1, 0.15, 0.3\}$, and for MFVI, the initial value for the learnable parameter variances has the most influence, and we test values $\in \{0.01815, 0.006715, 0.002476\}$. During inference time, 5 stochastic forward passes are used for MCD and MFVI to obtain prediction mean and variance. For RMDSRN to be trained with the exponential scheduler defined in Eq. (5) for the variance regularization, we test $\lambda_{max} \in \{10, 30, 100\}$ with $\lambda_{min} = 0$ and $r = 500$. In the following evaluation, we show results from the best trial among the three runs considering both reconstruction and variance quality. Please refer to the supplemental material for the quantitative results of all runs.

Data reconstruction evaluation. To quantify the reconstruction quality of uncertain SRNs, we compute the peak signal-to-noise ratio (PSNR) between the ground truth and the mean reconstructions from uncertain SRNs and present in Tab. 1. Results for non-uncertain SRNs are also included with the "SRN" model label. MDSRN demonstrates leading PSNR compared to Bayesian methods, DE, and PV in all datasets thanks to the feature-grid-sharing scheme that effectively scales up the member capacity. Tab. 1 shows MDSRN can outperform MCD, MFVI, and PV by up to +5 dB in Asteroid with NGP. Furthermore, the best RMDSRN among 3 training trials of different λ_{max} can further increase the PSNR slightly, making RMDSRN the most accurate uncertain SRN in the perspective of data-level PSNR. When comparing the accuracy of RMDSRN with the SRN without variance quantification, RMDSRN demonstrates competitive predictive performance close to the non-uncertain base SRN in all datasets except Plume. This indicates the potential of RMDSRN in providing quality reconstruction similar to the conventional non-uncertain models while being able to quantify its prediction confidence. A noteworthy observation for PV is that despite the close performance to RMDSRN in Nyx, Supernova, and Isotropic, it has the lowest PSNR for Plume and Asteroid. We observe that PV with NLL training can have a more noisy descent trajectory for the loss compared to other approaches, resulting in a more varying performance between experiments.

For visualization quality comparison, we provide volume rendering images of uncertain-SRN-reconstructed Asteroid and Plume in Fig. 4, and each image from SRNs is labeled with image-level metric results comparing against the ground truth including PSNR, structural similarity index measure (SSIM) [47], and Learned Perceptual Image Patch
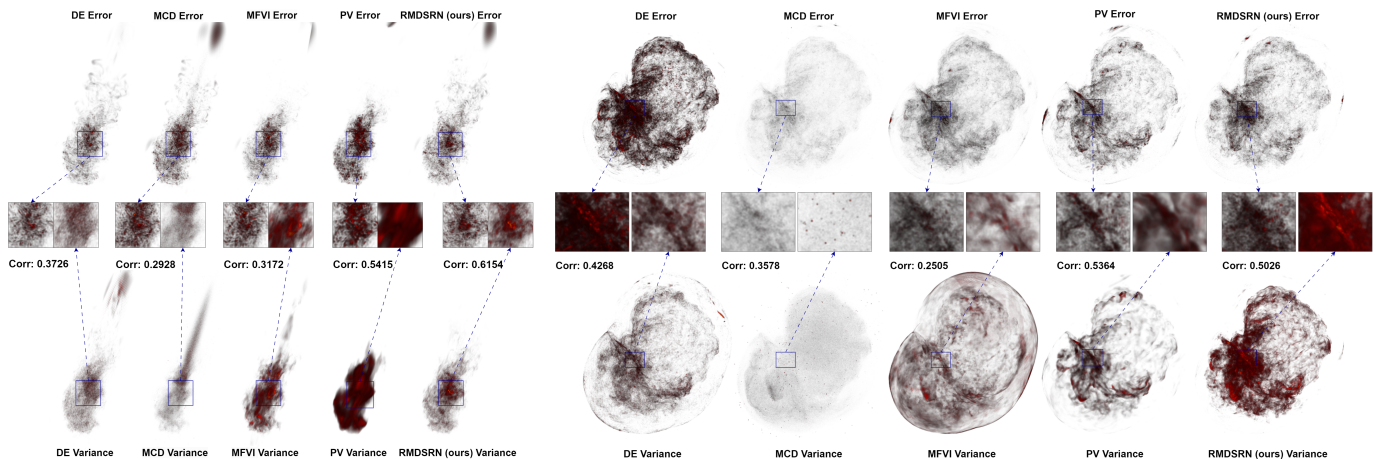
Fig. 5: Volume renderings of the highest error (top) versus variance (bottom) obtained from uncertain SRNs of Plume on the left and Supernova on the right. RMDSRN more precisely recovers fine-grained error patterns. While PV attains a high variance-error correlation as a strong competitor, the variance can be oversmooth, potentially overlooking detailed error structures.

Similarity (LPIPS) [54]. We omit MDSRN in the visual comparison since RMDSRN can represent the best quality of our methods, and the visual quality difference between their renderings is not significant. RMDSRN scores the best quantitative results with the lowest perceptual distance plus the highest image PSNR and SSIM, except for Asteroid where PV outperforms RMDSRN in PSNR. Nevertheless, with a scrutinization of the enlarged views in Asteroid, RMDSRN recovers the most detailed feature patterns compared to PV, which is consistent with the SSIM and LPIPS results. Similarly for Nyx, we can observe that RMDSRN preserves the shapes and colors of data features highlighted by the transfer function to the greatest extent, and its images look sharper with higher visual fidelity compared to the other renderings that appear fuzzier.

Variance quantification evaluation. To quantitatively evaluate the spatial similarity between the variance and error, we propose to use Pearson correlation (corr) and a modified Jaccard index with spatial tolerance (JI-ST) as the metrics. Given a variance field and an error field sampled with the same resolution as the data, a higher correlation value suggests high variances and errors are more likely to coincide in the same voxels and vice versa. Jaccard index [18] is computed as the ratio of intersection over union (IoU) between two sets of voxels with one having the highest variances and the other containing the highest errors. We propose a modified Jaccard index with spatial tolerance (JI-ST) to measure the overlap between the two voxel sets with top variances and errors. When calculating the intersection between the two sets, JI-ST expands the top-error voxel set with one-voxel neighbors of the original members. This relaxed intersection calculation considers that although the highest variances and errors do not always coincide in the same voxels, if the voxels in the variance set are sufficiently adjacent to those in the error set, the variances are still informative in exposing the weakness of the model in visualization. For JI-ST, we evaluate the voxel sets with top 1%- and 5%-ranked variances and errors. As the results in Tab. 1 show, PV and RMDSRN dominate the top one and two variance evaluation metrics. Although MDSRN does not demonstrate a clear advantage in the correlation and JI-ST, RMDSRN improves the similarity of variance to its error considerably and achieves competitive statistics for both metrics compared with PV, surpassing DE and BNNs by a great margin. Since the non-uncertain SRN does not provide variance quantification, the non-applicable metrics are shown as N/A.

Another uncertain SRN evaluation metric is NLL. The mean prediction and variance from the uncertain models are used to parameterize a Gaussian distribution of the prediction, and NLL evaluates how likely the ground truth data are generated from the predicted distributions with smaller values indicating more probable distributions. PV attains the best NLL in most datasets as expected, since NLL is directly minimized during training. After PV, we observe ensemble methods including RMDSRN and DE attain the highest quality distributions representing the data, despite MDSRN tending to produce less desirable Gaussians.
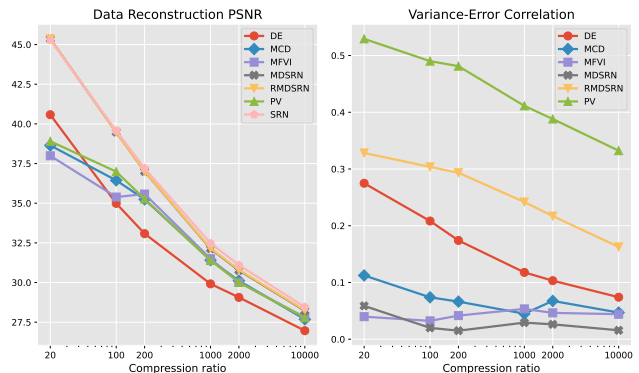

Fig. 6: Data reconstruction and variance-error correlation are evaluated for SRNs across six compression levels. MDSRN and RMDSRN lead the reconstruction PSNR in all model sizes, while PV demonstrates more advantageous Pearson correlations between variance and error.

To visually evaluate if the variance and error fields correlate well, we compare volume renderings of both fields for Plume (left) and Supernova (right) in Fig. 5. For each dataset, we visualize a percentage of voxels with the highest-ranked variances versus errors to study the effectiveness of variance in revealing regions with the most error-prone predictions. The data-space correlation is also included between each pair of the error and variance renderings. For Plume, qualitatively the variance of RMDSRN captures most precisely the high error regions shown in the cropped view. Although PV achieves a high variance-error correlation, the variance rendering demonstrates an excessive level of fuzziness that fails to reproduce the detailed error structures for Plume. As for Supernova, the fuzzy characteristic of the variance from PV is still present, but relatively more spatial structures are revealed than in Plume. The variance from RMDSRN displays an overall similar feature structure to the error with more fine-grained patterns captured than other methods.

Cross-compression-level evaluation. In addition to the above evaluations comparing uncertain SRNs under a constant size within each dataset and base SRN, we present quantitative results evaluating their performance under six compression ratios from 1/10000 to 1/20 for Isotropic with fV-SRN in Fig. 6. Similar to the setup for previous evaluations, we train 3 runs of MCD, MFVI, and RMDSRN with the mentioned hyperparameter sets and choose the best model with relatively better reconstruction and correlation. We plot the compression ratio versus reconstruction PSNR and variance-error correlation for applicable models. For the per-compression-level PSNR, we can observe RMDSRN and MDSRN in gray and yellow achieve leading PSNRs in every compression level, comparable to the SRN without variance

PSNR: 19.08, SSIM=0.8469, LPIPS=0.15908   PSNR: 31.27, SSIM=0.9534, LPIPS=0.07355   PSNR: 18.91, SSIM=0.8291, LPIPS=0.18457   PSNR: 35.82, SSIM=0.9701, LPIPS=0.05684   PSNR: 32.29, SSIM=0.9756, LPIPS=0.03981

Mean Field Rendering

Uncertainty- Aware Rendering

GT

PSNR: 32.78, SSIM=0.9552, LPIPS=0.08355   PSNR: 35.28, SSIM=0.9649, LPIPS=0.06469   PSNR: 32.09, SSIM=0.9318, LPIPS=0.11932   PSNR: 36.56, SSIM=0.9703, LPIPS=0.05924   PSNR: 40.06, SSIM=0.9855, LPIPS=0.02934

DE　　　　　　**MCD**　　　　　　**MFVI**　　　　　　**PV**　　　　　**RMDSRN (ours)**
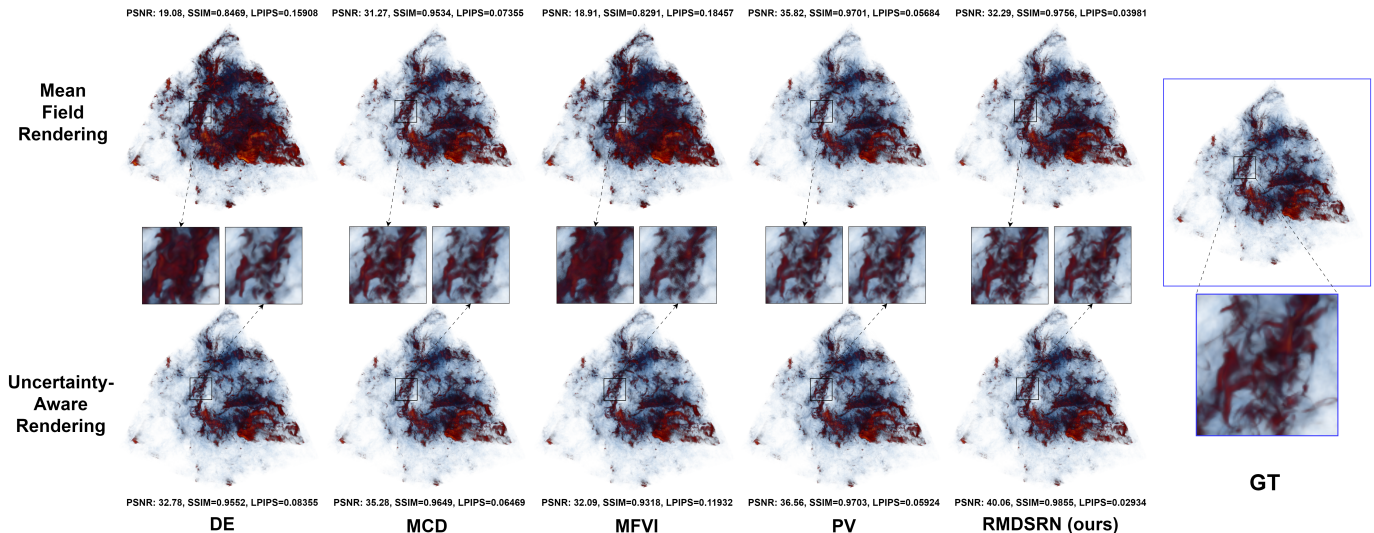
Fig. 7: Mean prediction volume rendering versus uncertainty-aware volume rendering with our adaptation of statistical DVR [33] for uncertain SRNs. All uncertain SRNs produce more accurate volume renderings with the uncertainty-aware algorithm than rendering the mean, revealing the potential advantage of uncertainty-aware visualization methods for uncertain SRNs.

estimation, and the accuracy gap for other uncertain SRNs increases with larger model sizes. Despite a potentially oversmooth variance as shown in the renderings from the variance evaluation, PV consistently outperforms other uncertain models with considerably higher variance-error correlations in the correlation chart. RMDSRN follows PV with the next best correlation scores. Notably, despite having inferior reconstruction accuracy, DE also exhibits high-quality variance across compression levels over Bayesian methods.

5.2 Uncertainty-Aware Volume Rendering with Uncertain SRNs

With uncertain SRNs outputting multiple predictions for any coordinate input, in addition to visualizing the variance as one presented approach for quality-informed visualization, the architectures open the opportunity for the application of probabilistic volume visualization algorithms that work on uncertain data to produce uncertainty-aware visualization [1, 8, 30–33].

We present our adaptations and results of the uncertainty-aware volume rendering technique proposed by Sakhaee et al. [33] for uncertain SRNs to directly incorporate the multiple predictions for every sampled location into the final rendered image, beyond rendering only the mean field. In the statistical direct volume rendering (DVR) framework from Sakhaee et al. [33], the uncertainty in the data values is integrated into transfer function (TF) classification. For uncertain data with the value in each spatial location conforming to some probability distribution, they propose to compute the expected color in a sampled location by applying the TF to all possible values and accumulating the colors as well as opacities weighted by the probabilities of the values. To adapt the statistical DVR framework for uncertain SRNs, we compute the expected color by the sum of colors of each predicted value weighted by the probability of the prediction under the Gaussian distribution parameterized by the mean value and variance. The weights need to be normalized by the total probabilities of all samples to ensure they sum to one before multiplying by the color and opacity. With this formulation, predictions closer to the mean are given higher weights than others. This behavior can be justified by the observation that the mean prediction is often more accurate than the samples for both ensemble and BNN methods evaluated, hence samples more similar to the mean are also expected to be higher-quality predictions, and they should contribute more to the expected color and opacity.

We render fV-SRN-based uncertain SRNs for Isotropic with the described adaptation of the uncertainty-aware TF classification method in comparison with the mean prediction renderings in Fig. 7. The models are from the experiments for Sec. 5.1. We observe the uncer-
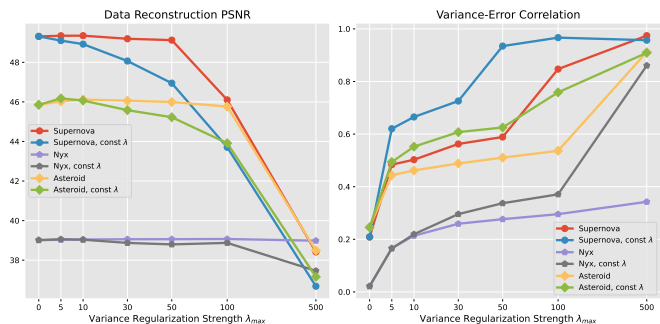

Fig. 8: Effect of varying variance regularization strength λ_{max} on RMDSRN with scheduler tested on Supernova, Nyx, and Asteroid. Results for RMDSRN with constant λ are shown with suffix "const λ". RMDSRN performs well under wide ranges of λ_{max} with the proposed scheduler compared to with a constant λ.

tain rendering approach improves the image accuracy considerably for all uncertain SRNs with RMDSRN showing the most accurate visualization for both mean prediction and uncertainty-aware rendering. By aggregating the post-classification colors instead of classifying the mean prediction with the TF, the features in red displayed in the enlarged views exhibit clearer visibility of the feature structure, and all uncertain SRNs attain higher image accuracy from the quantitative results reported besides the images.

5.3 Ablation Study: Regularization Strength and Scheduler

The variance regularization endows RMDSRN with a reliable variance closely correlated to the error down to detailed spatial patterns compared to other approaches, and this performance is tightly coupled with an appropriate regularization strength. We study the behavior of RMDSRN under different λ_{max} in Fig. 8 with $\lambda_{min} = 0$ and $r = 500$ as the settings of the λ scheduler, compared with an MDSRN baseline. We test on Supernova, Nyx, and Asteroid with network configurations for each dataset the same as in Sec. 5.1. In addition, we test RMDSRN with constant λ throughout training without the proposed scheduler, and the results are labeled with a suffix of "const λ".

Before we compare the performance of RMDSRN with a constant versus scheduled λ, we first focus on discussing the results for RMDSRN with the scheduler plotted in red, yellow, and purple lines in Fig. 8. The right subplot shows stronger regularization monotonically increases the variance-error correlation as shown in all datasets, whereas the reconstruction accuracy can degrade after some threshold.

Therefore, the ideal λ_{max} for the scheduler ranges from zero up to the maximum value before observing a significant accuracy drop for RMDSRN to achieve the best quality variance possible while retaining similar reconstruction accuracy as MDSRN. For Supernova with MDSRN scoring a PSNR of 49.31 dB, the best λ_{max} for RMDSRN is between 10 to 30 beyond which the accuracy starts to underperform MDSRN. This threshold is between 50-100 for Asteroid and beyond 500 for Nyx. Comparing the best λ_{max} between datasets with different levels of MDSRN accuracy, we can observe the threshold inversely correlates with model accuracy. In other words, for a given dataset and a base SRN architecture, RMDSRN with less capacity is expected to be more amenable to a high regularization strength than a larger RMDSRN that can learn a more accurate reconstruction.

Although the best-performing λ_{max} value can differ between training settings, it can be observed that a mild regularization with $\lambda_{max} \in [5, 10]$ works well generally, and there are only slight performance gaps comparing to the optimal λ_{max} due to diminishing returns evident in Fig. 8, hence we recommend this range as a default setting. If it is desired to maximize the benefit of the regularization, an MDSRN baseline can be trained to guide the λ_{max} tuning. Though the interpretation of PSNR values can vary for different datasets, we recommend trying λ_{max} up to 50 if the baseline MDSRN scores PSNR under 40 dB, or up to 30 for under 45 dB PSNR. As we have observed the inverse relationship between PSNR and λ_{max}, a more intelligent scheduler that automatically scales λ during training based on the current accuracy might be possible. We leave this to future work and recommend any RMDSRN to start with the suggested λ_{max} between 5 to 10.

We now compare the RMDSRNs trained with our scheduler versus constant strength, after we have established that a proper regularization should increase the variance quality without degrading the performance of RMDSRN under the MDSRN baseline. Taking Asteroid as an example, it is observed in the PSNR chart that RMDSRN with the scheduler can have λ_{max} set between 0 to 100 without hurting the predictive performance. However, with a constant λ, the results in the green line show the accuracy of RMDSRN already drops below MDSRN for strengths greater and equal to 30. The similar trend is manifested in other datasets as well. In particular, a constant λ of 5 is already excessive for Supernova. Concluding from the observations, our proposed λ scheduling routine is shown to ease the sensitivity of λ tuning, such that a wider range of strengths can be applied for better variance quantification over MDSRN without an adverse side effect on accuracy.

5.4 Ablation Study: Decoder Count

The number of decoder members can affect the performance of ensemble networks, and we study the performance of our MDSRN and RMDSRN with different numbers of member decoders while keeping the total model size constant. We test on Supernova with APGMSRN.

As Fig. 9 reveals, MDSRN does not exhibit an additional advantage in reconstruction accuracy with more decoders under a constant parameter budget. This adheres to our empirical observation reported in Sec. 4.1 that the most important factor affecting the ensemble accuracy is the capacity of member models. Nevertheless, the variance quality keeps improving from 3 to 15 decoders. Following the results, a decoder of 5 for MDSRN can be a proper choice balancing both accuracy and variance quality. For RMDSRN, the reconstruction PSNR shows a monotonic decrease for both λ_{max} of 5 and 10. Although RMDSRN with $\lambda_{max} = 30$ shows increased performance from 3 to 5, the lower performance of 3 decoders is likely a result of excessive regularization strength, as the accuracy is substantially lower than MDSRN. As for their variance quality, decoder count is less impactful as the regularization strength and the correlation change in response to different numbers of decoders is minimal. Concluding from the observations, 3 to 5 decoders for RMDSRN can provide a balanced performance considering accuracy and variance-error correlation.

6 LIMITATIONS AND FUTURE WORK

We acknowledge several limitations of our multi-decoder architecture and regularization with a discussion of possible improvements and
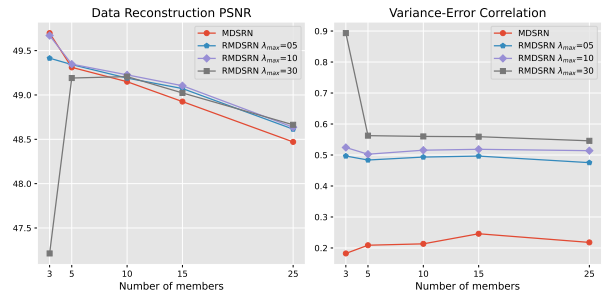

Fig. 9: Effect of varying number of decoders for MDSRN and RMDSRN with different variance regularization strengths.

future work. First of all, our MDSRN, as an ensemble method, requires training multiple member networks for each optimization step. Compared with non-ensemble methods such as BNNs and PV, MDSRN takes longer to train as shown in Tab. 1.

Secondly, although MDSRN achieves great parameter efficiency for better accuracy than DE, PV, and BNNs, the variance quality does not exhibit an advantage as shown in Tab. 1 without the variance regularization. The diversity of networks represented by the uncertain model is an important factor affecting the performance [7, 50], and improving the diversity of the decoders can be future work for better variance quality.

As studied in Sec. 5.3, regularization strength λ has important implications for the accuracy and variance quality of RMDSRN. The proposed λ scheduler is shown to reduce its sensitivity and provide a default setting that works for most experiments. However, λ_{max} still needs to be set, and finding the best λ_{max} can require hyperparameter searching. On the other hand, as the proper λ appears to inversely correlate with model accuracy, an automatic λ scheduler that adjusts the value based on both the training iteration and current prediction accuracy might be able to be developed as future work.

In this work, our utilization of the prediction variance is to facilitate error-aware data reconstruction and visualization, but it can be exploited for more diverse purposes. We plan to explore the potential of active learning methods with uncertain networks for resource-efficient training of SRNs for large-scale scientific datasets.

7 CONCLUSION

To equip feature grid SRNs for scientific data with inference time prediction quality assessment, we propose RMDSRN that provides prediction variance quantification from multiple plausible predictions made to any given input coordinate for error-aware reconstruction of volumetric data. RMDSRN comprises a parameter-efficient multi-decoder architecture synergized with a novel variance regularization loss for reliable variance estimation well-correlated with prediction error. Through both quantitative and qualitative evaluations, RMDSRN demonstrates superior data reconstruction and visualization accuracy as well as competitive variance quality under the same model size across different datasets and compression levels compared to alternative uncertain neural network architectures shown in Sec. 5.1. Additionally, we present results of uncertainty-aware volume rendering incorporated to uncertain SRNs, which reveals the potential advantage of leveraging uncertain predictions for a more accurate volume rendering than using the mean field with uncertain SRNs. As we examine the effect of regularization strength and decoder count on the performance of MDSRN and RMDSRN in the ablation studies, we find our models can achieve satisfactory reconstruction accuracy and variance quality with a relatively small number of members under 5, and our RMDSRN performs well under a wide range of regularization strengths thanks to the exponential growth scheduler derived for the regularization.

Our work explores an error-aware SRN method realized with uncertain neural network architecture, and we see the potential of continued research on more diverse error-aware SRN approaches and applications to improve the trustworthiness and performance of SRN for scientific visualization in future work.

REFERENCES

[1] 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. 8

[2] D. Bauer, Q. Wu, and K.-L. Ma. Fovolnet: Fast volume rendering using foveated deep neural networks. *IEEE transactions on visualization and computer graphics*, 29(1):515–525, 2022. 5

[3] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra. Weight uncertainty in neural network. In *International conference on machine learning*, pp. 1613–1622. PMLR, 2015. 1, 2, 3, 5

[4] L. Chen, W. Chen, R. Wang, and M. Pollefeys. Leveraging neural radiance fields for uncertainty-aware visual localization. *arXiv preprint arXiv:2310.06984*, 2023. 3

[5] Z. Eaton-Rosen, F. Bragman, S. Bisdas, S. Ourselin, and M. J. Cardoso. Towards safe deep learning: accurately quantifying biomarker uncertainty in neural network predictions. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pp. 691–699. Springer, 2018. 4

[6] F. Farokhmanesh, K. Höhlein, C. Neuhauser, T. Necker, M. Weissmann, T. Miyoshi, and R. Westermann. Neural Fields for Interactive Visualization of Statistical Dependencies in 3D Simulation Ensembles. In M. Guthe and T. Grosch, eds., *Vision, Modeling, and Visualization*. The Eurographics Association, 2023. doi: 10.2312/vmv.20231229 2

[7] S. Fort, H. Hu, and B. Lakshminarayanan. Deep ensembles: A loss landscape perspective. *arXiv preprint arXiv:1912.02757*, 2019. 9

[8] N. Fout and K.-L. Ma. Fuzzy volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2335–2344, 2012. 8

[9] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016. 1, 2, 3, 5

[10] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, et al. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56(Suppl 1):1513–1589, 2023. 2

[11] L. Goli, C. Reading, S. Sellán, A. Jacobson, and A. Tagliasacchi. Bayes' Rays: Uncertainty quantification in neural radiance fields. *arXiv preprint arXiv:2309.03185*, 2023. 3

[12] P. Gu, D. Z. Chen, and C. Wang. Nervi: Compressive neural representation of visualization images for communicating volume visualization results. *Computers & Graphics*, 116:216–227, 2023. 1, 2

[13] F. K. Gustafsson, M. Danelljan, and T. B. Schön. Evaluating scalable bayesian deep learning methods for robust computer vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, 2020. 4

[14] 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*, 2022. 2

[15] W. He, J. Wang, H. Guo, K.-C. Wang, H.-W. Shen, M. Raj, Y. S. 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, 2019. 5

[16] K. Höhlein, S. Weiss, T. Necker, M. Weissmann, T. Miyoshi, and R. Westermann. Evaluation of Volume Representation Networks for Meteorological Ensemble Compression. In J. Bender, M. Botsch, and D. A. Keim, eds., *Vision, Modeling, and Visualization*. The Eurographics Association, 2022. doi: 10.2312/vmv.20221198 1, 2

[17] E. Ilg, O. Cicek, S. Galesso, A. Klein, O. Makansi, F. Hutter, and T. Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 652–667, 2018. 4

[18] P. Jaccard. The distribution of the flora in the alpine zone. 1. *New phytologist*, 11(2):37–50, 1912. 7

[19] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, eds., *3rd International Conference on Learn-ing Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 6

[20] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017. 1, 2, 3, 4, 5

[21] O. Laurent, A. Lafage, E. Tartaglione, G. Daniel, J.-M. Martinez, A. Bursuc, and G. Franchi. Packed-Ensembles for Efficient Uncertainty Estimation. In *11th International Conference on Learning Representations*. Kigali, Rwanda, May 2023. 3

[22] I. Loshchilov and F. Hutter. SGDR: stochastic gradient descent with warm restarts. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. 5, 6

[23] Y. Lu, K. Jiang, J. A. Levine, and M. Berger. Compressive neural representations of volumetric scalar fields. In *Computer Graphics Forum*, vol. 40, pp. 135–146. Wiley Online Library, 2021. 1, 2, 3

[24] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2

[25] J. Mukhoti and Y. Gal. Evaluating bayesian deep learning methods for semantic segmentation. *arXiv preprint arXiv:1811.12709*, 2018. 4

[26] T. Müller, A. Evans, C. Schied, and A. Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 2, 3, 4, 6

[27] X. Pan, Z. Lai, S. Song, and G. Huang. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*, pp. 230–246. Springer, 2022. 3, 6

[28] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 6

[30] K. Pothkow and H.-C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1393–1406, 2010. 8

[31] K. Pöthkow and H.-C. Hege. Nonparametric models for uncertainty visualization. In *Computer Graphics Forum*, vol. 32, pp. 131–140. Wiley Online Library, 2013. 8

[32] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. In *Computer Graphics Forum*, vol. 30, pp. 931–940. Wiley Online Library, 2011. 1, 8

[33] E. Sakhaee and A. Entezari. A statistical direct volume rendering framework for visualization of uncertain data. *IEEE transactions on visualization and computer graphics*, 23(12):2509–2520, 2016. 1, 2, 5, 8

[34] V. Saragadam, J. Tan, G. Balakrishnan, R. G. Baraniuk, and A. Veeraraghavan. Miner: Multiscale implicit neural representation. In *European Conference on Computer Vision*, pp. 318–333. Springer, 2022. 2

[35] G. Scalia, C. A. Grambow, B. Pernici, Y.-P. Li, and W. H. Green. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. *Journal of chemical information and modeling*, 60(6):2697–2717, 2020. 4

[36] J. Shen, A. Agudo, F. Moreno-Noguer, and A. Ruiz. Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. In *European Conference on Computer Vision*, pp. 540–557. Springer, 2022. 3

[37] J. Shen, A. Ruiz, A. Agudo, and F. Moreno-Noguer. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*, pp. 972–981, 2021. doi: 10.1109/3DV53792.2021.00105 3

[38] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020. 2

[39] V. Sitzmann, M. Zollhöfer, and G. Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. In *Advances in Neural Information Processing Systems*, 2019. 2

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 3

[41] N. Sünderhauf, J. Abou-Chakra, and D. Miller. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In *2023*

IEEE International Conference on Robotics and Automation (ICRA), pp. 9370–9376. IEEE, 2023. 3

[42] M. Tancik, P. P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. T. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020. 2

[43] K. Tang and C. Wang. Ecnr: Efficient compressive neural representation of time-varying volumetric datasets. *arXiv preprint arXiv:2311.12831*, 2023. 1, 2

[44] M. O. Turkoglu, A. Becker, H. A. Gündüz, M. Rezaei, B. Bischl, R. C. Daudt, S. D'Aronco, J. Wegner, and K. Schindler. Film-ensemble: Probabilistic deep learning via feature-wise linear modulation. *Advances in Neural Information Processing Systems*, 35:22229–22242, 2022. 3

[45] M. Valdenegro-Toro. Deep sub-ensembles for fast uncertainty estimation in image classification. *arXiv preprint arXiv:1910.08168*, 2019. 3

[46] F. Vasconcelos, B. He, N. M. Singh, and Y. W. Teh. Uncertainr: Uncertainty quantification of end-to-end implicit neural representations for computed tomography. *Trans. Mach. Learn. Res.*, 2023, 2023. 3

[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[48] S. Weiss, P. Hermüller, and R. Westermann. Fast neural representations for direct volume rendering. In *Computer Graphics Forum*, vol. 41, pp. 196–211. Wiley Online Library, 2022. 1, 2, 3, 4, 6

[49] Y. Wen, D. Tran, and J. Ba. Batchensemble: an alternative approach to efficient ensemble and lifelong learning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. 3

[50] A. G. Wilson and P. Izmailov. Bayesian deep learning and a probabilistic perspective of generalization. *Advances in neural information processing systems*, 33:4697–4708, 2020. 9

[51] Q. Wu, D. Bauer, Y. Chen, and K.-L. Ma. Hyperinr: A fast and predictive hypernetwork for implicit neural representations via knowledge distillation. *arXiv preprint arXiv:2304.04188*, 2023. 2

[52] Q. Wu, D. Bauer, M. J. Doyle, and K.-L. Ma. Interactive volume visualization via multi-resolution hash encoding based neural representation. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 1, 2, 3, 4, 6

[53] S. W. Wurster, T. Xiong, H.-W. Shen, H. Guo, and T. Peterka. Adaptively placed multi-grid scene representation networks for large-scale data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 1, 2, 3, 4, 6

[54] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 7