

# On the Quantification of Image Reconstruction Uncertainty without Training Data

Jiaxin Zhang  
 Intuit AI Research  
 jxzhangai@gmail.com

Sirui Bi  
 Walmart Global Tech  
 siruijhu@gmail.com

Victor Fung  
 Georgia Institute of Technology  
 victorfung@gatech.edu

## Abstract

*Computational imaging plays a pivotal role in determining hidden information from sparse measurements. A robust inverse solver is crucial to fully characterize the uncertainty induced by these measurements, as it allows for the estimation of the complete posterior of unrecoverable targets. This, in turn, facilitates a probabilistic interpretation of observational data for decision-making. In this study, we propose a deep variational framework that leverages a deep generative model to learn an approximate posterior distribution to effectively quantify image reconstruction uncertainty without the need for training data. We parameterize the target posterior using a flow-based model and minimize their Kullback-Leibler (KL) divergence to achieve accurate uncertainty estimation. To bolster stability, we introduce a robust flow-based model with bi-directional regularization and enhance expressivity through gradient boosting. Additionally, we incorporate a space-filling design to achieve substantial variance reduction on both latent prior space and target posterior space. We validate our method on several benchmark tasks and two real-world applications, namely fastMRI and black hole image reconstruction. Our results indicate that our method provides reliable and high-quality image reconstruction with robust uncertainty estimation.*

## 1. Introduction

In computer vision and image processing, computational image reconstruction is a typical inverse problem where the goal is to learn and recover a hidden image  $\mathbf{x}$  from directly measured data  $\mathbf{y}$  via a forward operator  $\mathcal{F}$ . Such mapping  $\mathbf{y} = \mathcal{F}(\mathbf{x})$ , referred to as the *forward process* is often well-established. Unfortunately, the *inverse process*  $\mathbf{x} = \mathcal{F}^{-1}(\mathbf{y})$ , proceeds in the opposite direction, which is a nontrivial task since it is often ill-posed. A regularized optimization is formulated to recover the hidden image  $\mathbf{x}^*$ :

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \{ \mathcal{L}(\mathbf{y}, \mathcal{F}(\mathbf{x})) + \lambda \omega(\mathbf{x}) \}, \quad (1)$$

where  $\mathcal{L}$  is a loss function to measure the difference between the measurement data and the forward prediction,  $\omega$  is a

regularization function, and  $\lambda$  is a regularization weight. The regularization function, including  $\ell_1$ -norm and total variation, are often used to constrain the image to a unique inverse solution in under-sampled imaging systems [11, 39].

Recent trends have focused on using deep learning for computational image reconstruction, which does not rely on an explicit forward model or iterative updates but performs learned inversion from representative large datasets [7, 57], with applications in medical science, biology, astronomy and more. However, most of these existing studies in regularized optimization [32, 34] and feed-forward deep learning approaches [7, 42, 45] mainly focus on pursuing a unique inverse solution by recovering a single point estimate. This leads to a significant limitation when working with under-determined systems where it is conceivable that multiple inverse image solutions would be equally consistent with the measured data [4, 40]. Practically, in many cases, only partial and limited measurements are available which naturally leads to a *reconstruction uncertainty*. Thus, a reconstruction using a point estimate without uncertainty quantification would potentially mislead the decision-making process [55, 56]. Therefore, the ability to characterize and quantify reconstruction uncertainty is of paramount relevance. In principle, Bayesian methods are an attractive route to address the inverse problems with uncertainty estimation. However, in practice, the exact Bayesian treatment of complex problems is usually intractable [54]. The common limitation is to resort to inference and sampling, typically by Markov Chain Monte Carlo (MCMC), which are often prohibitively expensive for imaging problems due to the *curse of dimensionality*.

For nonlinear, non-convex image reconstruction problems, a deeper architecture with enough expressivity may be required to approximate the complex posterior distribution. However, increasing model depth will result in overfitting, as well as making sampling and computing inefficient [53]. The essential assumption of invertibility is also potentially violated along with instability issues caused by the aggregation of numerical errors. The imprecision and variation in flow-based models induce additional uncertainties. Moreover, the variation in posterior distribution caused by latent space sampling is also non-negligible for the evaluation of

the total reconstruction uncertainty. These challenges make accurate uncertainty estimation to be a nontrivial task.

**Main contributions.** We present a novel uncertainty-aware method for achieving reliable image reconstruction with an accurate estimation of data uncertainty resulting from measurement noise and sparsity. Our approach leverages a deep variational framework with robust generative flows and variance-reduced sampling to accurately characterize and quantify reconstruction uncertainty without any training data. We propose a flow-based variational approach to approximate the posterior distribution of a target image, minimizing model uncertainties by building a robust flow-based model with enhanced stability through bi-directional regularization and improved flexibility through gradient boosting. We conservatively propagate the statistics of latent distribution to the posterior through a deterministic invertible transformation, replacing simple random sampling with generalized Latin Hypercube Sampling to achieve significant variance reduction on posterior approximation. We demonstrate our method on fastMRI reconstruction and interferometric imaging problems, showing that it achieves reliable and high-quality reconstruction with accurate uncertainty evaluation.

## 2. Background

**Normalizing flows.** Generative models, such as GANs and VAEs, are intractable for explicitly learning the probability density function which plays a fundamental role in uncertainty estimation. Flow-based generative models overcome this difficulty with the help of *normalizing flows* (NFs), which describe the transformation from a latent density  $\mathbf{z}_0 \sim \pi_0(\mathbf{z}_0)$  to a target density  $\tau(\mathbf{x})$ , where  $\mathbf{x} = \mathbf{z}_K \sim \pi_K(\mathbf{z}_K)$  through a sequence of invertible mappings  $\mathcal{T}_k : \mathbb{R}^d \rightarrow \mathbb{R}^d, k = 1, \dots, K$ . By using the change of variables rule

$$\tau(\mathbf{x}) = \pi_k(\mathbf{z}_k) = \pi_{k-1}(\mathbf{z}_{k-1}) \left| \det \frac{\partial \mathcal{T}_k^{-1}}{\partial \mathbf{z}_{k-1}} \right|, \quad (2)$$

where the target density  $\pi_K(\mathbf{z}_K)$  obtained by successively transforming a random variable  $\mathbf{z}_0$  through a chain of  $K$  transformations  $\mathbf{z}_K = \mathcal{T}_K \circ \dots \circ \mathcal{T}_1(\mathbf{z}_0)$  is

$$\log \tau(\mathbf{x}) = \log \pi_K(\mathbf{z}_K) = \log \pi_0(\mathbf{z}_0) - \sum_{k=1}^K \log \left| \det \frac{\partial \mathcal{T}_k}{\partial \mathbf{z}_{k-1}} \right|$$

where each transformation  $\mathcal{T}_k$  must be sufficiently *expressive* while being theoretically *invertible* and efficient to compute the Jacobian determinant. Affine coupling functions [14, 24] are often used because they are simple and efficient to compute. However, these benefits come at the cost of expressivity and flexibility; many flows must be stacked to learn a complex representation.

**Density estimation.** Assuming that samples  $\{\mathbf{x}_i\}_{i=1}^M$  drawn from a probability density  $p(\mathbf{x})$  are available, our goal is to learn a flow-based model  $\tau_\phi(\mathbf{x})$  parameterized by the vector  $\phi$  through a transformation  $\mathbf{x} = \mathcal{T}(\mathbf{z})$  of a latent density  $\pi_0(\mathbf{z})$  with  $\mathcal{T} = \mathcal{T}_K \circ \dots \circ \mathcal{T}_1$  as a  $K$ -step flow. This is achieved by minimizing the KL-divergence  $D_{\text{KL}} = \text{KL}(p(\mathbf{x}) \parallel \tau_\phi(\mathbf{x}))$ , which is equivalent to maximum likelihood estimate (MLE).

**Variational inference.** The goal is to approximate the posterior distribution  $p$  through a variational distribution  $\pi_K$  encoded by a flow-based model  $\tau_\phi(\mathbf{x})$ , which is tractable to compute and draw samples. This is achieved by minimizing the KL-divergence  $D_{\text{KL}} = \text{KL}(\pi_K \parallel p)$ , which is equivalent to maximizing an evidence lower bound

$$\mathcal{V}_\phi(\mathbf{x}) = \mathbb{E}_{\mathbf{z}_K \sim \pi_K(\mathbf{z}_K)} [-\log p(\mathbf{x}, \mathbf{z}_K) + \log \pi_K(\mathbf{z}_K | \mathbf{x})].$$

**Evaluation metrics for deep generative models.** Designing indicative evaluation metrics for generative models and samples remains a challenge. A commonly used metric for measuring the similarity between real and generated images has been the Fréchet Inception Distance (FID) score [20] but it fails to separate two critical aspects of the quality of generative models: *fidelity* that refers to the degree to which the generated samples resemble the real ones, and *diversity*, which measures whether the generated samples cover the full variability of the real samples. Sajjadi *et al.* [36] proposed the two-value metrics (*precision* and *recall*) to capture the two characteristics separately. Recently, Naeem *et al.* [31] introduced two reliable metrics (*density* and *coverage*) to evaluate the quality of the generated posterior samples and measure the difference between them and ground truth. *density* improves upon the precision metric by dealing with the overestimation issue and *coverage* instead of recall metric is to better measure the diversity by building the nearest neighbor manifolds around the true samples.

## 3. Methodology

### 3.1. Deep Variational Framework

Our goal is to build a deep variational framework to accurately estimate the data uncertainty quantified by an approximation of the posterior distribution. The regularized optimization for solving inverse problems can be written in terms of data fidelity (data fitting loss) and regularity:

$$\begin{aligned} \mathbf{x}^* &= \arg \min_{\mathbf{x}} \{ \mathcal{L}_D(\mathbf{y}, \mathcal{F}(\mathbf{x})) + \lambda \omega(\mathbf{x}) \} \\ &= \arg \min_{\mathbf{x}} \left\{ \underbrace{\| \mathbf{y} - \mathcal{F}(\mathbf{x}) \|^2}_{\text{Data fidelity}} + \underbrace{\lambda \omega(\mathbf{x})}_{\text{Regularity}} \right\}. \end{aligned} \quad (3)$$

Assuming the forward operator  $\mathcal{F}$  is known and the measurement noise statistics are given, we can reformulate the inverse problem in a probabilistic way. In the Bayesian

perspective, the regularized inverse problem in Eq. 3 can be solved by Bayesian inference but aims to maximize the posterior distribution by searching a point estimator  $\mathbf{x}^*$ :

$$\begin{aligned} \mathbf{x}^* &= \arg \max_{\mathbf{x}} \underbrace{\left\{ \log p(\mathbf{x}|\mathbf{y}) \right\}}_{\text{Posterior}} \\ &= \arg \max_{\mathbf{x}} \left\{ \underbrace{\log p(\mathbf{y}|\mathbf{x})}_{\text{Data likelihood}} + \underbrace{\log p(\mathbf{x})}_{\text{Prior}} \right\}, \end{aligned} \quad (4)$$

where the prior distribution  $p(\mathbf{x})$  (e.g., image prior [42] in reconstruction problems) defines a similar regularization term and data likelihood  $p(\mathbf{y}|\mathbf{x})$  corresponds to the data fidelity in Eq. 3. If we parameterize the target  $\mathbf{x}$  using a generative model  $\mathbf{x} = \mathcal{T}_\phi(\mathbf{z}), \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  with model parameter  $\phi$ , an approximate posterior distribution  $\tau_{\phi^*}(\mathbf{x})$  is obtained by minimizing the KL-divergence between the generative distribution and the target posterior distribution

$$\begin{aligned} \phi^* &= \arg \min_{\phi} \text{KL}(\tau_\phi(\mathbf{x}) \parallel p(\mathbf{x}|\mathbf{y})) \\ &= \arg \min_{\phi} \mathbb{E}_{\mathbf{x} \sim \tau_\phi(\mathbf{x})} [-\log p(\mathbf{y}|\mathbf{x}) \\ &\quad - \log p(\mathbf{x}) + \log \tau_\phi(\mathbf{x})]. \end{aligned} \quad (5)$$

Unfortunately, the probability density (likelihood)  $\tau_\phi(\mathbf{x})$  cannot be exactly evaluated by most existing generative models, such as GANs or VAEs. Flow-based models offer a promising approach to computing the likelihood exactly via the change of variable theorem with invertible architectures. Therefore, we can reformulate the equation in terms of a flow-based model as

$$\begin{aligned} \phi^* &= \arg \min_{\phi} \mathbb{E}_{\mathbf{z} \sim \pi(\mathbf{z})} [-\log p(\mathbf{y}|\mathcal{T}_\phi(\mathbf{z})) \\ &\quad - \log p(\mathcal{T}_\phi(\mathbf{z})) + \log \pi(\mathbf{z}) - \log |\det \nabla_{\mathbf{z}} \mathcal{T}_\phi(\mathbf{z})|]. \end{aligned} \quad (6)$$

Replacing data likelihood and prior terms by using data fidelity loss and regularization function in Eq. 3, we can define a new optimization problem where it can be approximated by a Monte Carlo method in practice:

$$\begin{aligned} \phi^* &= \arg \min_{\phi} \mathbb{E}_{\mathbf{z} \sim \pi(\mathbf{z})} [\mathcal{L}_D(\mathbf{y}, \mathcal{F}(\mathcal{T}_\phi(\mathbf{z}))) + \lambda\omega(\mathcal{T}_\phi(\mathbf{z})) \\ &\quad + \log \pi(\mathbf{z}) - \log |\det \nabla_{\mathbf{z}} \mathcal{T}_\phi(\mathbf{z})|] \\ &= \arg \min_{\phi} \sum_{j=1}^M \left[ \mathcal{L}_D(\mathbf{y}, \mathcal{F}(\mathcal{T}_\phi(\mathbf{z}_j))) + \lambda\omega(\mathcal{T}_\phi(\mathbf{z}_j)) \right. \\ &\quad \left. - \underbrace{\log |\det \nabla_{\mathbf{z}} \mathcal{T}_\phi(\mathbf{z}_j)|}_{\text{Entropy}} \right], \end{aligned} \quad (7)$$

where  $\pi(\mathbf{z})$  is a constant and  $\log |\det \nabla_{\mathbf{z}} \mathcal{T}_\phi(\mathbf{z}_j)|$  is entropy that is critical to encourage sample diversity and exploration to avoid generative models from collapsing to a deterministic solution.

Note that the flow-based model is very critical and sensitive to uncertainty estimation and quantification within this variational framework. To perform accurate data uncertainty estimation, the uncertainty associated with the flow-based model must be minimized. To this end, we develop a Robust Generative Flow (RGF) model with enhanced stability, expressivity, and flexibility, while conserving efficient inference and sampling without increasing architecture depth.

## 3.2. Robust Generative Flows (RGF)

### 3.2.1 Stability of invertible architectures

The proposed variational framework relies on the essential assumption of the theoretical invertibility in flow-based models. However, this assumption is challenged by recent studies [5, 6, 25] with findings that the commonly used invertible architectures  $\mathcal{T}_\phi$ , such as additive and affine coupling blocks, suffer from exploding inverses and thus prone to becoming numerically non-invertible [21], which will violate the assumption underlying their main advantages, including efficient sampling and exact likelihood estimation. Typically, the coupling blocks show an exploding inverse effect because the singular values of the forward mapping tend to zero as depth increases. The numerical errors introduced in both forward and inverse mapping will aggravate the imprecision and instability which renders the architecture non-invertible and results in uncontrollable model uncertainties that are intractable to characterize.

The stability of invertible neural network (INN) architectures can be analyzed by the property of bi-Lipschitz continuity if there exists a constant  $L := \text{Lip}(\mathcal{T})$  and a constant  $L^* := \text{Lip}(\mathcal{T}^{-1})$  such that for all  $x_1, x_2; y_1, y_2 \in \mathbb{R}^d$

$$\begin{aligned} \|\mathcal{T}(x_1) - \mathcal{T}(x_2)\| &\leq L\|x_1 - x_2\|, \\ \|\mathcal{T}^{-1}(y_1) - \mathcal{T}^{-1}(y_2)\| &\leq L^*\|y_1 - y_2\|. \end{aligned} \quad (8)$$

Penalty terms on the Jacobian can be used to enforce INN stability locally. If  $\mathcal{T}$  is Lipschitz continuous and differentiable, we have  $\text{Lip}(\mathcal{T}) = \sup_{\mathbf{x} \in \mathbb{R}^d} \|J_{\mathcal{T}}(\mathbf{x})\|_2$ . Instead of estimating  $\text{Lip}(\mathcal{T})$  using random samples, we use finite differences (FD) to approximate  $\text{Lip}(\mathcal{T})$  as

$$\begin{aligned} \text{Lip}(\mathcal{T}) &= \sup_{\mathbf{x} \in \mathbb{R}^d} \|J_{\mathcal{T}}(\mathbf{x})\|_2 \\ &\approx \sup_{\mathbf{x} \in \mathbb{R}^d} \sup_{\|\nu\|_2=1} \frac{1}{\varepsilon} \|\mathcal{T}(\mathbf{x}) - \mathcal{T}(\mathbf{x} + \varepsilon\nu)\|_2, \end{aligned} \quad (9)$$

where  $\varepsilon > 0$  is a step size in FD. This penalty term (as a regularizer) can be added to the loss function on both directions  $\mathcal{T}$  and  $\mathcal{T}^{-1}$  such that we have a stable forward and inverse mapping. This bi-directional FD regularization can remedy the non-invertible failures in many coupling blocks including spline function.

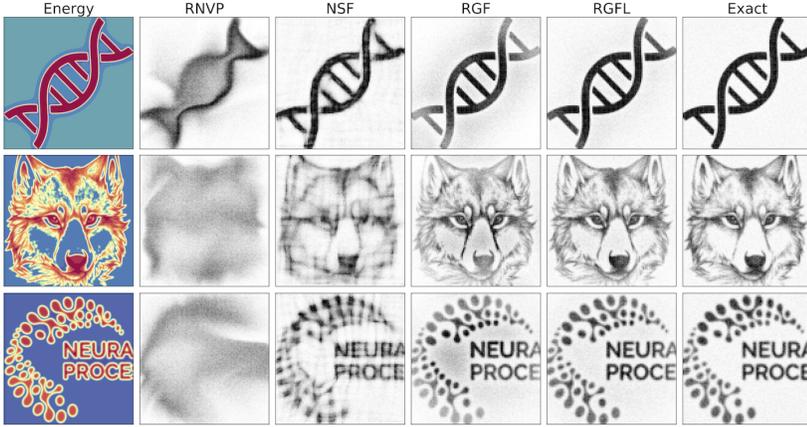


Figure 1. Sampling of 2D densities using variational inference with energy function.

Metrics	RNVP	NSF	RGF	RGFL
Precision↑	0.951	0.988	0.992	<b>0.996</b>
Recall↑	0.996	0.994	<b>0.997</b>	<b>0.997</b>
Density↑	0.823	0.962	0.987	<b>0.989</b>
Coverage↑	0.909	0.926	0.950	<b>0.961</b>
Metrics	RNVP	NSF	RGF	RGFL
Precision↑	0.988	0.993	0.994	<b>0.997</b>
Recall↑	0.997	0.998	0.998	<b>0.999</b>
Density↑	0.889	0.948	0.989	<b>1.001</b>
Coverage↑	0.912	0.951	0.946	<b>0.964</b>
Metrics	RNVP	NSF	RGF	RGFL
Precision↑	0.969	0.987	0.993	<b>0.994</b>
Recall↑	0.994	<b>0.996</b>	<b>0.996</b>	<b>0.996</b>
Density↑	0.829	0.942	0.987	<b>1.003</b>
Coverage↑	0.877	0.934	0.940	<b>0.965</b>

Table 1. Evaluation metrics.

### 3.2.2 Enhanced expressivity and flexibility

Recent trends in NFs have focused on creating deeper, more complex transformations to increase the flexibility of the learned distribution. However, a deeper structure of flows often renders instability and uncertainty caused by aggregation of the numerical errors. Also, with greater model complexity comes a greater risk of overfitting while slowing down training, sampling, and inference [17]. To address these issues, we propose to use a gradient boosting approach for increasing the expressiveness of the neural spline flow (NSF) model [15]. Our new model is built by iteratively adding new NF components with gradient boosting, where each new NF component is fit to the residual of the previously trained components. A weight is then learned via stochastic gradient descent for each component, which results in a mixture model structure, whose flexibility increases as more components are added.

**Gradient boosting flow model.** A gradient boosting flow model is constructed by successively adding new components, where each new component  $t_K^{(c)}$  is a K-step normalizing flow that matches the functional gradient of the loss function from the  $(c-1)$  previously trained flow components  $\mathcal{T}_K^{(c-1)}$ . Typically, the gradient boost flow model is constructed by a convex combination of fixed and new flow components:

$$\mathcal{T}_K^{(c)}(\mathbf{z} | \mathbf{x}) = (1 - \beta_c)\mathcal{T}_K^{(c-1)}(\mathbf{z} | \mathbf{x}) + \beta_c t_K^{(c)}(\mathbf{z} | \mathbf{x}), \quad (10)$$

where  $\mathbf{x}$  are the observed data,  $\mathbf{z}$  are the latent variables and  $\beta_c$  is the weight to the new component where  $\beta_c \in [0, 1]$  to sure the mixture model in Eq. 10 is a valid probability distribution.

To pursue a variational posterior that closely matches the true posterior, which corresponds to the reverse KL-divergence  $\text{KL}(\tau_\phi(\mathbf{x}) \parallel p(\mathbf{x} | \mathbf{y}))$ . Thus, we seek to minimize

the variational bound:

$$\mathcal{V}_\phi(\mathbf{x}) = \mathbb{E}_{\mathcal{T}_K^{(c)}} [\log \mathcal{T}_K^{(c)}(\mathbf{z}_K | \mathbf{x}) - \log p(\mathbf{y} | \mathbf{x}, \mathbf{z}_K)]. \quad (11)$$

**Boosting components updating.** Given the objective function in Eq. 11, we proceed with deriving updates for new boosting components. First, at the current stage  $c$ , we let  $\mathcal{T}_K^{(c-1)}$  to be fixed, and the target is learning the component  $t_K^{(c)}$  and the weight  $\beta_c$  based on functional gradient descent (FGD) [30]. We take the gradient of objective in Eq. 11 with respect to  $\mathcal{T}_K^{(c)}$  at  $\beta_c \rightarrow 0$ :

$$\nabla_{\mathcal{T}_K^{(c)}} \mathcal{V}_\phi(\mathbf{x}) \Big|_{\beta_c \rightarrow 0} = -\log \frac{p(\mathbf{y} | \mathbf{x}, \mathbf{z}_K)}{\mathcal{T}_K^{(c-1)}(\mathbf{z} | \mathbf{x})}. \quad (12)$$

Since  $\mathcal{T}_K^{(c-1)}(\mathbf{z} | \mathbf{x})$  are the fixed components, then minimizing the loss  $\mathcal{V}_\phi(\mathbf{x})$  can be achieved by selecting a new component  $t_K^{(c)}$  that has the maximum inner product with the negative of the gradient [30]. As a result, we can choose a  $t_K^{(c)}(\mathbf{z} | \mathbf{x})$  based on:

$$t_K^{(c)}(\mathbf{z} | \mathbf{x}) = \arg \min_{t_K \in \mathcal{T}_K} \mathbb{E}_{t_K(\mathbf{z} | \mathbf{x})} \left[ -\log \frac{p(\mathbf{y} | \mathbf{x}, \mathbf{z}_K^{(c)})}{\mathcal{T}_K^{(c-1)}(\mathbf{z}_K^{(c)} | \mathbf{x})} \right] \quad (13)$$

where  $\mathbf{z}_K^c$  denotes a sample transformed by component  $c$ 's flow.

**Components weights updating.** Once the  $t_K^{(c)}(\mathbf{z}_K | \mathbf{x})$  is estimated, the gradient boost flow model needs to determine the corresponding  $\beta_c \in [0, 1]$ . However, jointly optimizing both  $t_K^{(c)}$  and the weights  $\beta_c$  is a challenging optimization problem. We therefore consider a two-step optimization strategy which means that we first train  $t_K^{(c)}$  until convergence and then optimize the corresponding weight  $\beta_c$  by using the objective in Eq. 13. Similarly, the weights on each

component can be updated by using the gradient of the loss  $\mathcal{V}_\phi(\mathbf{x})$  with respect to  $\beta_c$ , as shown in Algorithm 1.

$$\frac{\partial \mathcal{V}_\phi(\mathbf{x})}{\partial \beta_c} = \sum_{i=1}^n \left( \mathbb{E}_{t_K^{(c)}(\mathbf{z}|\mathbf{x}_i)} \left[ \xi_{\beta_c}^{(s-1)}(\mathbf{z}|\mathbf{x}_i) \right] - \mathbb{E}_{\mathcal{T}_K^{(c-1)}(\mathbf{z}|\mathbf{x}_i)} \left[ \xi_{\beta_c}^{(s-1)}(\mathbf{z}|\mathbf{x}_i) \right] \right) \quad (14)$$

where  $\xi_{\beta_c}^{(s-1)}(\mathbf{z}|\mathbf{x}_i)$  is defined as  $\xi_{\beta_c}^{(s-1)}(\mathbf{z}|\mathbf{x}_i) = \log \mathcal{T}_K^{(c-1)}(\mathbf{z}|\mathbf{x}) - \log p(\mathbf{y}|\mathbf{x}, \mathbf{z}_K)$ .

---

#### Algorithm 1 Updating component weights $\beta_c$ by SGD

---

- 1: Requirement and initialization: predefined  $C$ , step size  $\lambda$ , tolerance  $\epsilon$ . Set  $s = 0$ ,  $\beta_c^{(0)} = 1/C$
  - 2: **While**  $|\beta_c^{(s)} - \beta_c^{(s-1)}| < \epsilon$  **do**
  - 3: Generate samples  $\mathbf{z}_{K,i}^{(c-1)} \sim \mathcal{T}_K^{(c-1)}(\mathbf{z}|\mathbf{x}_i)$  and  $\mathbf{z}_{K,i}^{(c)} \sim t_K^{(c)}(\mathbf{z}|\mathbf{x}_i)$  for  $i = 1, \dots, n$
  - 4: Estimate gradients using the MC method  $\nabla_{\beta_c} \mathcal{V}_\phi(\mathbf{x}) = 1/n \sum_{i=1}^n \left[ \xi_{\beta_c}^{(s-1)}(\mathbf{z}_{K,i}^{(c)}|\mathbf{x}_i) - \xi_{\beta_c}^{(s-1)}(\mathbf{z}_{K,i}^{(c-1)}|\mathbf{x}_i) \right]$
  - 5: Update weights  $\beta_c^{(s)} = \beta_c^{(s-1)} - \lambda \nabla_{\beta_c}$  and clip weights  $\beta_c^{(s)}$  to  $[0, 1]$
  - 6:  $s = s + 1$
  - 7: **return**  $\beta_c^{(s)}$
- 

### 3.3. Variance-reduced Latent Sampling

Rather than using simple random sampling (SRS) with a larger variance, Latin Hypercube Sampling (LHS) is an ideal candidate with variance reduction, which is generalized in terms of a spectrum of stratified sampling, referred to as partially stratified sampling (PSS), which shows to reduce variance associated with variable interaction but LHS reduces variance associated with additive (main) effects. A hybrid combination of LHS and PSS, named Latinized partially stratified sampling (LPSS) proposed by [37] can reduce variance associated with variable interaction and additive effects simultaneously. However, classical LHS used in LPSS suffers from a lack of exploratory capability due to its random pair scheme. To better explore all possible inverse solutions, we propose to leverage maximin criteria - maximizes the minimum distance between all pairs of points,

$$X_n = \arg \max \min \{d(\mathbf{x}_i, \mathbf{x}_k) : i \neq k = 1, \dots, M\} \quad (15)$$

where  $d(\mathbf{x}, \mathbf{x}') = \sum_{k=1}^M (\mathbf{x}_i - \mathbf{x}'_k)^2$  is the Euclidean distance. Instead of random pair, these criteria will greatly improve the space-filling properties of LPSS, specifically in high-dimensional space. Note that, unlike quasi-Monte Carlo (QMC) methods [12], e.g., Sobol sequence shown in Fig. 2 as well, which are limited in high dimensional problem

[27], but LPSS performs well with space-filling property and variance reduction in high dimensions.

Algorithm 2 shows the workflow of our proposed variational framework with robust generative flows and variance-reduced latent sampling given a single measurement data.

---

#### Algorithm 2 RGF in variational perspective

---

- 1: **Requirements:** RGF model  $\mathcal{T}_\phi$  parameterized by  $\phi$ , number of gradient boosting (GB) component  $C$ , FD regularization coefficient  $\lambda$ , forward operator  $\mathcal{F}$ , measurement data  $\mathbf{y}$ , training batch size  $b_z$ , evaluation batch size  $e_z$ ,
  - 2: Draw random samples from the latent space  $\mathbf{z}_j \sim p(\mathbf{z})$  using LPSS, where the sample size is  $b_z \times i_z$
  - 3: Generate image samples by  $\mathbf{x}_j = \mathcal{T}_\phi(\mathbf{z}_j)$  where  $\mathcal{T}_\phi(\mathbf{z}_j)$  is defined by initial gradient boosting (GB) mixture model in Eq. 10 with bi-directional FD regularized loss in Eq. 9.
  - 4: Predict the measurement  $\mathbf{y}$  by evaluating the forward operator  $\mathbf{y}_j = \mathcal{F}(\mathbf{x}_j) = \mathcal{F}(\mathcal{T}_\phi(\mathbf{z}_j))$
  - 5: Evaluate the total loss  $\mathcal{L}_{\text{total}}$  defined in Eq. 7 by a summation of the data fidelity, prior and entropy loss.
  - 6: Update new gradient boosting components based on Eq. 13 with a two-stage optimization strategy
  - 7: Update boosting component weights according to Eq. 14 with stochastic gradient descent in Algorithm 1.
  - 8: After training is done, generate posterior samples, and estimate the statistical quantities.
- 

### 3.4. RGF Representational Capability

Next, we demonstrate that our proposed RGF can improve the representational capability of deterministic NFs at a given network size. Here we use images to define complicated two-dimensional distributions as the target distributions to be sampled. This benchmark inspired by [51] aims at generating high-quality images from the exact density to compare the performance of different generative models with four blocks. Fig.1 shows the sampling of DNA, Fox, and NeurIPS logo cases with different methods. As expected, RealNVP [14] has limitations in representational capability, which results in a blurred image without detailed structures. NSF [15] outperforms RNVP with a better structure on 2D images, but still fails to resolve details, specifically in the Fox and NeurIPS logo cases, at the selected neural architecture. Note that all the flow-based models tend to have a better representational capability as the depth increases but we fix their depth in four blocks to perform a fair comparison.

Our proposed RGF (with SRS) achieves high-quality approximation through a combination of gradient boosting and stable regularization with a shallow architecture. Although the samples in RGF are close to the ground truth, their performance can be further improved by using LPSS. Although

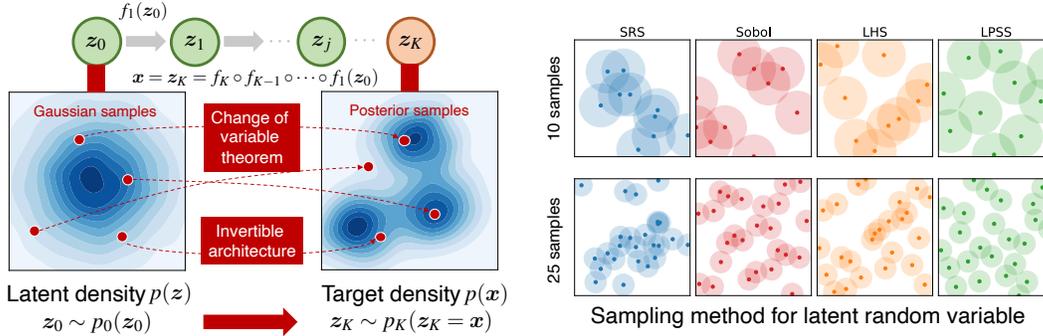


Figure 2. (Left) Illustration of mechanism for normalizing flows. The prior Gaussian samples drawn from the latent distribution are transformed to the posterior samples that match the target density by using a sequence of invertible mappings. (Right) Illustration of prior samples using four sampling algorithms for latent variables.

visual differences may be slight between RGF and RGFL, we differ them with a quantitative comparison using evaluation metrics (see Table 1). RGFL shows superior performance on both fidelity and diversity aspects of the generated images. In other words, the RGF model boosted by LPSS achieves not only a more accurate approximation to the true samples but also better captures the uncertainty (variation) in real samples. Both advantages are useful for quantifying the reconstruction uncertainty.

## 4. Experiments

### 4.1. Experimental setup

**Tasks and datasets.** We evaluate our method on two image reconstruction tasks: (1) compressed sensing MRI, from fastMRI dataset [52] with images of size  $320 \times 320$ ; and (2) Interferometric imaging, using blackhole images from the Event Horizon Telescope (EHT) [1, 2], with images of size  $160 \times 160$ . For both cases, we only use **one data**, which contains **one** specific measurement and the corresponding ground truth. The original images are resized to  $128 \times 128$  for fastMRI and  $32 \times 32$  for blackhole cases.

**Implementation.** We use 6 flow steps and 4 residual blocks for each step so that we have a total of 24 transformations in our RGF model. The whole model is trained on a single V100 GPU by using 20000 epochs with a batch size of 32 and an initial learning rate of  $1E-4$  in Adam.

**Evaluation metrics.** After training, 1000 samples drawn from the learned posterior distribution are used to evaluate the *mean*, *standard deviation*, and *absolute error*, which represents the bias between the mean image and ground truth. To evaluate the performance of the generative model, we provide a quantitative evaluation of the sample fidelity using *precision* and *density*, and the sample diversity using *recall* and *coverage*.

**Baselines.** We compare our proposed methods with a couple of methods including (1) conditional variational autoencoder

(cVAE) [38], which is typically used as a baseline method; (2) Deep Probabilistic Imaging (DPI), proposed by [40], mainly uses the RealNVP model for image Reconstruction; (3) GlowIP, which uses invertible generative models for inverse problems and image reconstruction [3]; (4) NSF, which is neural spline flows [15] used in our deep variational framework. RGFL is our proposed method which uses robust generative flows (RGF) with Latinized partially stratified sampling (LPSS). Although some recent methods [45, 48, 49], look very promising, we can not compare them if there are no available open-source codes.

Metrics	Brain case with $4\times$ speedup				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	8.87E-5	1.21E-5	8.82E-6	6.90E-6	<b>9.73E-7</b>
Abs. Error ↓	8.72E-5	3.31E-5	6.91E-6	1.83E-6	<b>1.78E-6</b>
	Brain case with $6\times$ speedup				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	1.02E-4	1.97E-5	1.02E-5	7.33E-6	<b>2.35E-6</b>
Abs. Error ↓	1.39E-4	5.70E-5	8.04E-6	6.99E-6	<b>4.58E-6</b>
	Brain case with $8\times$ speedup				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	2.58E-4	3.83E-5	2.27E-5	9.14E-6	<b>4.10E-6</b>
Abs. Error ↓	4.65E-4	8.81E-5	1.33E-5	8.04E-6	<b>7.12E-6</b>

Table 2. FastMRI brain case results compared to baseline methods.

### 4.2. FastMRI Case Study

Partial and under-sampled noisy measurements in MRI will lead to reconstruction uncertainty. We show that our RGFL method can be successfully applied to quantify the reconstruction variation and bias on two cases (brain and knee) with acceleration factors  $4\times$ ,  $6\times$ , and  $8\times$ . Fig. 3 shows the reconstruction results with pixel-wise statistics of the estimated posterior distribution. For both cases with a speedup  $4\times$  factor, our RGFL shows a more accurate mean estimate with a smaller absolute error than the DPI baseline with a larger architecture. Our advantage in terms of standard

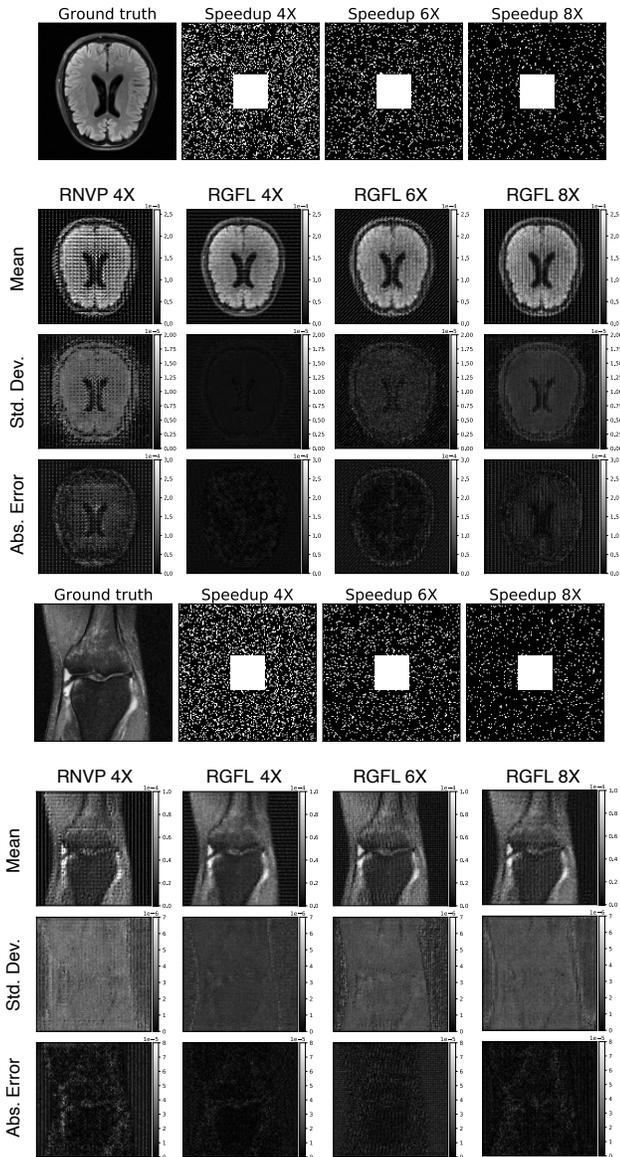


Figure 3. FastMRI reconstruction of brain case (up) and knee case (bottom) at three different acceleration speedup factors: 4X, 6X, and 8X (each shown in a column). Row 1 shows the ground truth and sampling masks for each case. Row 2-4 shows the mean, standard deviation, and absolute error for the estimated posterior samples.

deviation is more significant due to higher model robustness and variance reduction. Although the pixel-wise variance of the reconstruction tends to be larger as the speedup factor increases, our method provides a more reliable image reconstruction compared with other baselines.

We further use the mean of the standard deviation and the absolute error to quantitatively compare the pixel-wise statistics (see Table 2 and 3). Our method outperforms the other baselines in terms of accuracy (bias) and variation of the reconstruction. Specifically, our estimation achieves sig-

Metrics	Knee case with 4× speedup				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	7.98E-6	5.94E-6	4.33E-6	3.84E-6	<b>1.24E-6</b>
Abs. Error ↓	7.99E-6	4.83E-6	2.55E-6	9.36E-7	<b>7.03E-7</b>
Metrics	Knee case with 6× speedup				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	9.53E-6	9.33E-6	6.48E-6	4.71E-6	<b>2.05E-6</b>
Abs. Error ↓	1.02E-5	6.02E-6	4.87E-6	1.96E-6	<b>9.11E-7</b>
Metrics	Knee case with 8× speedup				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	3.61E-5	1.60E-5	9.96E-6	5.18E-6	<b>3.71E-6</b>
Abs. Error ↓	3.45E-5	7.94E-6	6.07E-6	2.06E-6	<b>1.15E-6</b>

Table 3. FastMRI knee case results compared to baseline methods.

nificant variance reduction with 1-2 orders of magnitude. Regarding the sample fidelity and diversity metrics, our method also shows competitive performance in most cases.

### 4.3. Interferometric Imaging Case Study

Our approach can be also applied to study radio interferometric astronomical imaging which was used to take the first black hole images. Sparse spatial frequency measurements are used to recover the underlying astronomical image (see Fig. 4). Relatively large telescope-based gain and phase error in the measurement noise lead to a non-convex image reconstruction problem where a challenge is the potential for multi-modal posterior distribution — different solutions fit the same measurement data visually and reasonably well. In this case, a synthetic black hole is used to illustrate our capability on reconstructed uncertainty estimation and multiple modes detection [40].

Metrics	Reconstructed black hole images (mode 1)				
	cVAE	DPI	NSF	GrowIP	RGFL
Std. Dev. ↓	1.23E-3	6.09E-4	5.43E-4	2.17E-4	<b>1.20E-4</b>
Abs. Error ↓	9.76E-4	3.85E-4	2.65E-4	<b>1.06E-4</b>	1.08E-4

Table 4. Statistical comparison of the estimated posteriors on the black hole image reconstruction (mode 1).

Fig. 4 shows multiple samples drawn from the learned generative model. Note that two different modes of reconstruction are captured by our RGFL model. t-SNE plots present a clear clustering of samples into two modes. Mode 1 fits the target image better than mode 2 which exhibits roughly 180-degree rotations. We perform statistical analysis for each mode to quantify the mean and standard deviation of the multi-modal posterior. Note that the uncertainty characterized by standard deviation shows a similar shape to the mean estimator. Mode 1, which is close to the correct solution, results in a smaller bias estimator than mode 2 (see Table 4). In this case, the space-filling sampling would be

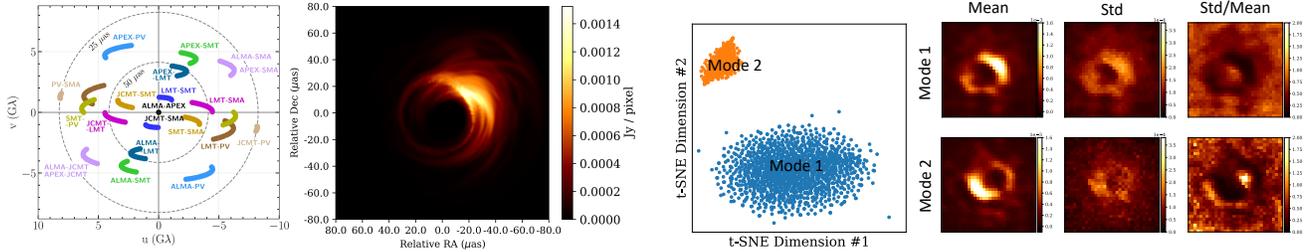


Figure 4. (Left) Measured frequency samples for EHT observing the M87\* [1, 2] and target synthetic black hole image; (Right) t-SNE result of posterior samples and statistics of the posterior samples on each model [40].

more important to capture the multi-modal posterior distribution. This can also be observed by the recall and coverage metrics in Table 4. Our method outperforms the other baselines with more accurate and reliable results.

## 5. Related work

**Bayesian deep learning.** Deep learning for solving inverse problems requires uncertainty estimation to be reliable in real settings. Bayesian deep learning [22, 23, 50], specifically Bayesian neural networks [19] can achieve this goal while offering a computationally tractable way for recovering reconstruction uncertainty. However, exact inference in the BNN framework is not a trivial task, so several variational approximation approaches are proposed to deal with the scalability challenges. Monte Carlo dropout [16] can be seen as a promising alternative approach that is easy to implement and evaluate. Deep ensemble [28] methods proposed by combining multiple deep models from different initializations have outperformed BNN. Recent methods on deterministic uncertainty quantification [43, 44] use a single forward pass but scale well to large datasets. Although these approaches show impressive performance, they rely on supervised learning with paired input-output datasets and only characterize the uncertainty conditioned on a training set.

**Variational approaches.** Variational methods offer a more efficient alternative approximating true but intractable posterior distribution by an optimally selected tractable distribution family [9]. However, the restriction to limited distribution families fails if the true posterior is too complex. Recent advances in conditional generative models, such as conditional GANs (cGANs) [46], overcome this restriction in principle, but have limitations in satisfactory diversity in practice. Another commonly adopted option is conditional VAEs (cVAEs) [38], which outperform cGANs in some cases, but in fact, the direct application of both conditional generative models in computational imaging is challenging because a large number of data is typically required [41]. This introduces additional difficulties if our observations and measurements are expensive to collect.

**Deep flow-based models.** Many very recent efforts have been made to solve inverse problems via deep generative

models [3, 8, 10, 13, 40, 49]. The flow-based model is a potential alternative via learning of a nonlinear transformation between the true posterior distribution and a simple prior distribution [14, 18, 24, 33, 35, 51]. These flow-based models possess critical properties: (a) the neural architecture is *invertible*, (b) the forward and inverse mapping is efficiently *computable* and (c) the Jacobian is *tractable*, which allows explicit computation of posterior probabilities. Fully invertible neural networks (INNs) are a natural choice to satisfy these properties and can be built using coupling layers, as introduced in the RealNVP [14] which is simple and efficient to compute. Although in principle, RealNVP layers are theoretically invertible, the actual computational of their inverse is not stable and sometimes even non-invertible due to aggravating numerical errors [6]. Many efforts have been spent to improve the stability, invertibility, flexibility, and expressivity of the flow-based models [26, 33, 47, 51], which inspired us to extend for the task of computing posterior in real-world imaging reconstruction.

Recently, Asim *et al.* [3] focused on producing a point estimate motivated by the MAP formulation and Wang *et al.* [49] aims at studying the full distributional recovery via variational inference. A follow-up study from [48] is to study image inverse problems with a normalizing flow prior, which is achieved by proposing a formulation that views the solution as the maximum a posteriori estimate of the image conditioned on the measurements. Our work is motivated by these recent advances but focuses on how to assess the image reconstruction (data) uncertainty with an explicitly known forward model with very sparse observations.

## 6. Conclusion

We propose an uncertainty-aware framework that leverages a deep variational approach with robust generative flows and variance-reduced sampling to perform an accurate estimation of image reconstruction uncertainty. The results on multiple benchmarks and real-world tasks demonstrate our advantages in uncertainty estimation. Although the current RGFL methods show superior performance on these imaging reconstruction tasks, the total computational cost will be a major concern if we plan to scale to complex high-resolution images, e.g., large-scale scientific simulations [29].

## References

- [1] Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. First m87 event horizon telescope results. ii. array and instrumentation. *The Astrophysical Journal Letters*, 875(1):L2, 2019.
- [2] Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. First m87 event horizon telescope results. iii. data processing and calibration. *The Astrophysical Journal Letters*, 875(1):L3, 2019.
- [3] Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. Invertible generative models for inverse problems: mitigating representation error and dataset bias. In *International Conference on Machine Learning*, pages 399–409. PMLR, 2020.
- [4] Riccardo Barbano, Željko Kereta, Chen Zhang, Andreas Hauptmann, Simon Arridge, and Bangti Jin. Quantifying sources of uncertainty in deep learning-based image reconstruction. *arXiv preprint arXiv:2011.08413*, 2020.
- [5] Jens Behrmann, Will Grathwohl, Ricky TQ Chen, David Duvenaud, and Jörn-Henrik Jacobsen. Invertible residual networks. In *International Conference on Machine Learning*, pages 573–582. PMLR, 2019.
- [6] Jens Behrmann, Paul Vicol, Kuan-Chieh Wang, Roger Grosse, and Jörn-Henrik Jacobsen. Understanding and mitigating exploding inverses in invertible neural networks. In *International Conference on Artificial Intelligence and Statistics*, pages 1792–1800. PMLR, 2021.
- [7] Chinmay Belthangady and Loic A Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature methods*, 16(12):1215–1225, 2019.
- [8] Sirui Bi, Victor Fung, and Jiaxin Zhang. Accelerating inverse learning via intelligent localization with exploratory sampling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14711–14719, 2023.
- [9] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.
- [10] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International Conference on Machine Learning*, pages 537–546. PMLR, 2017.
- [11] Charles Bouman and Ken Sauer. A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on image processing*, 2(3):296–310, 1993.
- [12] Russel E Caflisch et al. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 1998:1–49, 1998.
- [13] Giannis Daras, Joseph Dean, Ajil Jalal, and Alexandros G Dimakis. Intermediate layer optimization for inverse problems using deep generative models. *arXiv preprint arXiv:2102.07364*, 2021.
- [14] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.
- [15] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. *arXiv preprint arXiv:1906.04032*, 2019.
- [16] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.
- [17] Robert Gjaquinto and Arindam Banerjee. Gradient boosted normalizing flows. *Advances in Neural Information Processing Systems*, 33, 2020.
- [18] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.
- [19] José Miguel Hernández-Lobato and Ryan Adams. Probabilistic backpropagation for scalable learning of bayesian neural networks. In *International Conference on Machine Learning*, pages 1861–1869. PMLR, 2015.
- [20] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6629–6640, 2017.
- [21] Judy Hoffman, Daniel A Roberts, and Sho Yaida. Robust learning with jacobian regularization. *arXiv preprint arXiv:1908.02729*, 2019.
- [22] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NIPS*, 2017.
- [23] Mohammad Khan, Didrik Nielsen, Voot Tangkaratt, Wu Lin, Yarin Gal, and Akash Srivastava. Fast and scalable bayesian deep learning by weight-perturbation in adam. In *International Conference on Machine Learning*, pages 2611–2620. PMLR, 2018.
- [24] Diederik P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *arXiv preprint arXiv:1807.03039*, 2018.
- [25] Ivan Kobyzev, Simon Prince, and Marcus Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [26] Zhifeng Kong and Kamalika Chaudhuri. The expressive power of a class of normalizing flow models. In *International Conference on Artificial Intelligence and Statistics*, pages 3599–3609. PMLR, 2020.
- [27] Sergei Kucherenko, Daniel Albrecht, and Andrea Saltelli. Exploring multi-dimensional spaces: A comparison of latin hypercube and quasi monte carlo sampling techniques. *arXiv preprint arXiv:1505.02350*, 2015.
- [28] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NIPS*, 2017.
- [29] Alexander Lavin, David Krakauer, Hector Zenil, Justin Gottschlich, Tim Mattson, Johann Brehmer, Anima Anandkumar, Sanjay Choudry, Kamil Rocki, Atılım Güneş Baydin, et al. Simulation intelligence: Towards a new generation of scientific methods. *arXiv preprint arXiv:2112.03235*, 2021.

- [30] Llew Mason, Jonathan Baxter, Peter Bartlett, and Marcus Frean. Boosting algorithms as gradient descent. *Advances in neural information processing systems*, 12, 1999.
- [31] Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning*, pages 7176–7185. PMLR, 2020.
- [32] Frank Natterer and Frank Wübbeling. *Mathematical methods in image reconstruction*. SIAM, 2001.
- [33] Didrik Nielsen, Priyank Jaini, Emiel Hooeboom, Ole Winther, and Max Welling. Survae flows: Surjections to bridge the gap between vaes and flows. *Advances in Neural Information Processing Systems*, 33, 2020.
- [34] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.
- [35] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pages 1530–1538. PMLR, 2015.
- [36] Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 5234–5243, 2018.
- [37] Michael D Shields and Jiaxin Zhang. The generalization of latin hypercube sampling. *Reliability Engineering & System Safety*, 148:96–108, 2016.
- [38] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems*, 28:3483–3491, 2015.
- [39] David Strong and Tony Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165, 2003.
- [40] He Sun and Katherine L Bouman. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. *arXiv preprint arXiv:2010.14462*, 2020.
- [41] Francesco Tonolini, Jack Radford, Alex Turpin, Daniele Facio, and Roderick Murray-Smith. Variational inference for computational imaging inverse problems. *Journal of Machine Learning Research*, 21(179):1–46, 2020.
- [42] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.
- [43] Joost van Amersfoort, Lewis Smith, Andrew Jesson, Oscar Key, and Yarin Gal. Improving deterministic uncertainty estimation in deep learning for classification and regression. *arXiv preprint arXiv:2102.11409*, 2021.
- [44] Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In *International Conference on Machine Learning*, pages 9690–9700. PMLR, 2020.
- [45] Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 2(12):737–748, 2020.
- [46] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.
- [47] Yu Wang, Ján Drgoňa, Jiaxin Zhang, Karthik Somayaji Nangud Suryanarayana, Malachi Schram, Frank Liu, and Peng Li. Autof: Automated architecture optimization of normalizing flows with unconstrained continuous relaxation admitting optimal discrete solution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 10244–10252, 2023.
- [48] Jay Whang, Qi Lei, and Alex Dimakis. Solving inverse problems with a flow-based noise model. In *International Conference on Machine Learning*, pages 11146–11157. PMLR, 2021.
- [49] Jay Whang, Erik Lindgren, and Alex Dimakis. Composing normalizing flows for inverse problems. In *International Conference on Machine Learning*, pages 11158–11169. PMLR, 2021.
- [50] Andrew Gordon Wilson and Pavel Izmailov. Bayesian deep learning and a probabilistic perspective of generalization. *arXiv preprint arXiv:2002.08791*, 2020.
- [51] Hao Wu, Jonas Köhler, and Frank Noé. Stochastic normalizing flows. *arXiv preprint arXiv:2002.06707*, 2020.
- [52] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An open dataset and benchmarks for accelerated MRI. 2018.
- [53] Guannan Zhang, Jiaxin Zhang, and Jacob Hinkle. Learning nonlinear level sets for dimensionality reduction in function approximation. *Advances in Neural Information Processing Systems*, 32, 2019.
- [54] Jiaxin Zhang. Modern monte carlo methods for efficient uncertainty quantification and propagation: A survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 13(5):e1539, 2021.
- [55] Zizhao Zhang, Adriana Romero, Matthew J Muckley, Pascal Vincent, Lin Yang, and Michal Drozdal. Reducing uncertainty in undersampled mri reconstruction with active acquisition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2049–2058, 2019.
- [56] Qingping Zhou, Tengchao Yu, Xiaoqun Zhang, and Jinglai Li. Bayesian inference and uncertainty quantification for medical image reconstruction with poisson data. *SIAM Journal on Imaging Sciences*, 13(1):29–52, 2020.
- [57] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.