

Deep Probabilistic Imaging: Uncertainty Quantification and Multi-modal Solution Characterization for Computational Imaging

He Sun, Katherine L. Bouman,

California Institute of Technology
1200 E California Blvd, Pasadena, California 91125
hesun@caltech.edu, klbouman@caltech.edu

Abstract

Computational image reconstruction algorithms generally produce a single image without any measure of uncertainty or confidence. Regularized Maximum Likelihood (RML) and feed-forward deep learning approaches for inverse problems typically focus on recovering a point estimate. This is a serious limitation when working with underdetermined imaging systems, where it is conceivable that multiple image modes would be consistent with the measured data. Characterizing the space of probable images that explain the observational data is therefore crucial. In this paper, we propose a variational deep probabilistic imaging approach to quantify reconstruction uncertainty. Deep Probabilistic Imaging (DPI) employs an untrained deep generative model to estimate a posterior distribution of an unobserved image. This approach does not require any training data; instead, it optimizes the weights of a neural network to generate image samples that fit a particular measurement dataset. Once the network weights have been learned, the posterior distribution can be efficiently sampled. We demonstrate this approach in the context of interferometric radio imaging, which is used for black hole imaging with the Event Horizon Telescope, and compressed sensing Magnetic Resonance Imaging (MRI).

Introduction

Uncertainty quantification and multi-modal solution characterization are essential for analyzing the results of underdetermined imaging systems. In computational imaging, reconstruction methods solve an inverse problem to recover a hidden image from measured data. When this inverse problem is ill-posed there are infinite image solutions that fit the observed data. Occasionally these varied images lead to different scientific interpretations; thus it is important to be able to characterize the distribution of possibilities. In a Bayesian framework, this problem could ideally be addressed by accurately modeling the measurement noise, formulating an estimation problem, and computing the posterior distribution of the hidden image. However, this analytical approach is only tractable in simple cases. When the inverse problem is non-convex or the measurement noise is complicated (e.g., non-Gaussian) the posterior distribution can quickly become intractable to compute analytically.

Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

In this paper, we propose *Deep Probabilistic Imaging* (DPI), a new variational inference approach that employs a deep generative model to learn a reconstructed image’s *posterior distribution*. More specifically, we represent the posterior probability distribution using an invertible flow-based generative model. By training with a traditional maximum *a posteriori* (MAP) loss, along with a loss that encourages distribution entropy, the network converges to a generative model that approximately samples from the image posterior; the model returns a posterior sample image as output for a random Gaussian sample input. Our proposed approach enables uncertainty quantification and multi-modal solution characterization in non-convex inverse imaging problems. We demonstrate our method on the applications of compressed sensing MRI and astronomical radio interferometric imaging. High resolution radio interferometric imaging often requires a highly non-convex forward model, occasionally leading to multi-modal solutions.

Related Work

Computational Imaging

The goal of a computational image reconstruction method is to recover a hidden image from measured data. Imaging systems are often represented by a deterministic forward model, $y = f(x)$, where y are the observed measurements of a hidden image, x . A regularized maximum likelihood (RML) optimization can be solved to reconstruct the image:

$$\hat{x} = \arg \min_x \{\mathcal{L}(y, f(x)) + \lambda \mathcal{R}(x)\}, \quad (1)$$

where \hat{x} is the reconstructed image, $\mathcal{L}(\cdot, \cdot)$ is the data fitting loss between the measured data and the forward model prediction, $\mathcal{R}(\cdot)$ is a regularization function on the image, and λ is the hyper-parameter balancing the data fitting loss and the regularization term. The regularization function is necessary for obtaining a unique solution in under-constrained systems. Commonly used regularization functions include L1 norm, total variation (TV) or total squared variation (TSV) (Bouman and Sauer 1993; Kuramochi et al. 2018), maximum entropy (MEM) (Skilling and Bryan 1984), multivariate Gaussian (Zoran and Weiss 2011), and sparsity in the wavelet transformed domain (Candes and Romberg 2007).

Assuming the forward model and measurement noise statistics are known, one can derive the probability of measurements, y , conditioned on the hidden image, x , as $p(y|x)$.

In a Bayesian estimation framework, the regularized inverse problem can be interpreted as a maximum a posteriori (MAP) estimation problem:

$$\hat{x} = \arg \max_x \{\log p(y|x) + \log p(x)\}, \quad (2)$$

where the log-likelihood of the measurements correspond to the negative data fitting loss in Eq. 1, and the prior distribution of the image defines a regularization function (also referred to as an image prior).

Deep Image Prior In lieu of constraining the image via an explicit regularization function, $\mathcal{R}(\cdot)$, Deep Image Prior approaches (Ulyanov, Vedaldi, and Lempitsky 2018) parameterize the image as a deep neural network and implicitly force the image to respect low-level image statistics via the structure of a neural network:

$$w^* = \arg \min_w \mathcal{L}(y, f(g_w(z))) \quad \text{where } \hat{x} = g_{w^*}(z). \quad (3)$$

In this approach, $g_w(\cdot)$ is a deep generator, w are the neural network weights, w^* are the optimized weights, and z is a randomly chosen, but fixed, hidden state. The network weights are randomly initialized and optimized using gradient descent. This approach has been demonstrated in many applications, including inpainting, denoising and phase retrieval (Ulyanov, Vedaldi, and Lempitsky 2018; Bostan et al. 2020). Note that since the hidden state, z , is fixed, only a single reconstructed image is derived after optimization.

Uncertainty Quantification

Uncertainty quantification is important for understanding the confidence of recovered solutions in inverse problems.

MCMC Sampling One widely used approach for quantifying the uncertainty in image reconstruction is the Markov Chain Monte Carlo (MCMC) method. MCMC provides a way to approximate the posterior distribution of a hidden image via sampling (Bardsley 2012; Broderick et al. 2020). However, the MCMC approach can be prohibitively slow for high dimensional inverse problems (Cai, Pereyra, and McEwen 2018).

Bayesian Hypothesis Testing Uncertainty quantification can also be formulated as a Bayesian hypothesis test in linear imaging inverse problems (Repetti, Pereyra, and Wiaux 2019). In a linear imaging problem, the data likelihood is a log-concave function, which makes the hypothesis test on a specific image structure an efficient convex optimization program. However, this method cannot be generalized to imaging inverse problems, when either the negative data likelihood or the regularizer is non-convex.

Variational Bayesian Methods Variational inference is widely used to approximate intractable posterior distributions (Blei, Kucukelbir, and McAuliffe 2017; Arras et al. 2019). Instead of directly computing the exact posterior distribution, variational Bayesian methods posit a simple family of density functions and solve an optimization problem to find a member of this family closest to the target posterior distribution. Variational Bayesian methods are much faster than sampling methods (e.g., MCMC), and typically

achieve comparable performance (Gershman, Hoffman, and Blei 2012). The performance of variational methods depends on the modeling capacity of the density function family and the complexity of the target distribution. A commonly used technique for simplifying the density function is the mean-field approximation (Jordan et al. 1999), where the distributions of each hidden variables are assumed independent. The density function can also be parameterized using neural networks, such as flow-based generative models (Rezende and Mohamed 2015).

Bayesian Neural Networks Deep learning has become a powerful tool for computational imaging reconstruction (Ongie et al. 2020). Most current deep learning approaches only focus on the point estimate of the hidden image; however, Bayesian neural networks (MacKay 1995; Gal 2016) have been developed to quantify the uncertainty of deep learning image reconstructions (Xue et al. 2019). In particular, the weights of a Bayesian network are modeled as probability distributions, so that different predictions are obtained every time the network is executed. Although these approaches can achieve impressive performance, they rely on supervised learning and only describe the reconstruction uncertainty conditioned on a training set.

Empirical Sampling An alternative empirical approach for obtaining multiple candidate reconstructions is to solve a regularized inverse problem (Eq. 1) multiple times with different choices of regularizer hyper-parameters (e.g. λ in Eq. 1) and image initializations. This approach was used in (The EHT Collaboration et al. 2019) to characterize the uncertainty of the reconstructed black hole image M87*. Although the mean and standard deviation of these images provide a measure of uncertainty, there is no expectation that these samples satisfy a posterior distribution defined by the measurement data. In fact, this method only quantifies the reconstruction uncertainty due to choices in the reconstruction methods, such as regularizer hyper-parameters, instead of the uncertainty due to measurement noise and sparsity.

Flow-based Generative Models

Flow-based generative models are a class of generative models used in machine learning and computer vision for probability density estimation. These models approximate an arbitrary probability distribution by learning an invertible transformation of a generic distribution $\pi(z)$ (e.g. normal distribution). Mathematically, a flow-based generative model is represented as

$$x = G_\theta(z), \quad z = G_\theta^{-1}(x), \quad (4)$$

where $G_\theta(\cdot)$ is an invertible deep neural network parameterized by θ that links a sample from the target distribution, x , with a hidden state, z . The application of invertible transformations enables both efficient sampling, as well as exact log-likelihood computation. The log-likelihood of a sample can be analytically computed based on the ‘‘change of variables theorem’’:

$$\log q_\theta(x) = \log \pi(z) - \log \left| \det \frac{dG_\theta(z)}{dz} \right|, \quad (5)$$

where $\det \frac{dG_\theta(z)}{dz}$ is the determinant of the generative model's Jacobian matrix.

To keep the computation of the second term tractable, the neural network function, $G_\theta(\cdot)$, is restricted to forms such as NICE(Dinh, Krueger, and Bengio 2014), Real-NVP(Dinh, Sohl-Dickstein, and Bengio 2016) and Glow(Kingma and Dhariwal 2018). In these network architectures, the Jacobian matrix is a multiplication of only lower triangular matrices or quasi-diagonal matrices, which leads to efficient computation of the determinant.

Method

In this paper, we propose a variational Bayesian method to learn an approximate posterior distribution for the purpose of efficiently characterizing uncertainty in underdetermined imaging systems. We parameterize the latent image distribution using a flow-based generative model,

$$x \sim q_\theta(x) \Leftrightarrow x = G_\theta(z), z \sim \mathcal{N}(0, 1) \quad (6)$$

and learn the model's weights by minimizing the Kullback-Leibler (KL) divergence between the generative model distribution, $q(x)$, and the image posterior distribution, $p(x|y) \propto p(y|x)p(x)$:

$$\begin{aligned} \theta^* &= \arg \min_{\theta} D_{\text{KL}}(q_\theta(x) \| p(x|y)) \quad (7) \\ &= \arg \min_{\theta} \int q_\theta(x) [\log q_\theta(x) - \log p(x|y)] dx \\ &= \arg \min_{\theta} \int q_\theta(x) [\log q_\theta(x) - \log p(y|x) - \log p(x)] dx \\ &= \arg \min_{\theta} \mathbb{E}_{x \sim q_\theta(x)} [-\log p(y|x) - \log p(x) + \log q_\theta(x)]. \end{aligned}$$

Note that this loss can be interpreted as an expectation over the maximum *a posteriori* (MAP) loss from Eq. 2 with an added term to encourage entropy on the image distribution. Minimizing the negative entropy term, $H_\theta = \mathbb{E}_{x \sim q_\theta(x)} [\log q_\theta(x)]$, prevents the generative model from collapsing to a deterministic solution.

For most deep generative models the sample likelihood, $q_\theta(x)$, cannot be evaluated exactly. However, since a sample's likelihood can be computed according to Eq. 5 for flow-based models, this stochastic optimization problem can be rewritten as

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{z \sim \mathcal{N}(0,1)} \left\{ -\log p(y|G_\theta(z)) - \log p(G_\theta(z)) + \log \pi(z) - \log \left| \det \frac{dG_\theta(z)}{dz} \right| \right\}. \quad (8)$$

Approximating the expectation using a Monte Carlo method, and replacing the data likelihood and prior terms with the data fitting loss and regularization functions from Eq. 1, we obtain the optimization problem

$$\theta^* = \arg \min_{\theta} \sum_{k=1}^N \left\{ \mathcal{L}(y, f(G_\theta(z_k))) + \lambda \mathcal{R}(G_\theta(z_k)) - \log \left| \det \frac{dG_\theta(z_k)}{dz_k} \right| \right\}, \quad (9)$$

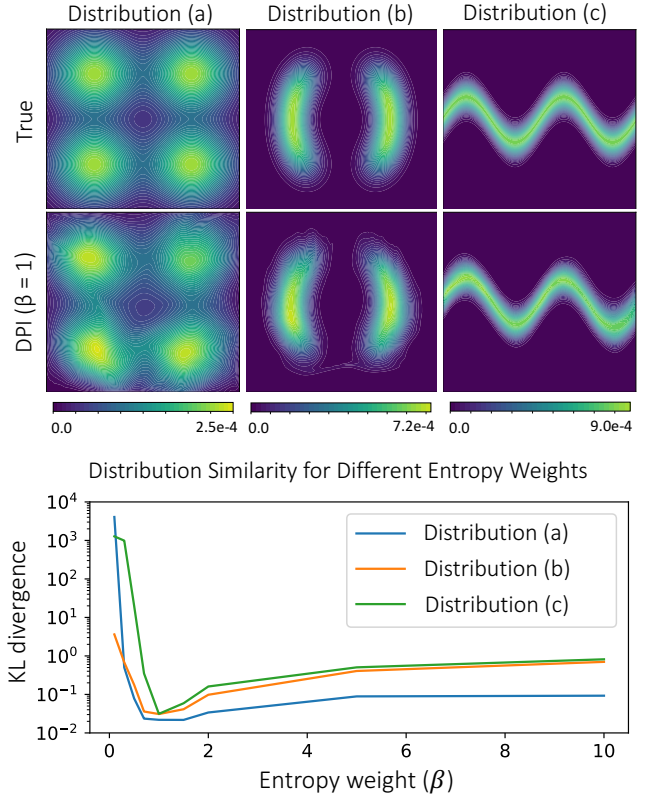


Figure 1: (Top) Three two-dimensional posterior distributions and their learned DPI approximations for entropy weight $\beta = 1$. Contours are overlaid to highlight differences between the true and learned distributions. (Bottom) Similarity between the true and learned distributions under different entropy weights β . As expected, $\beta = 1$ minimizes the KL divergence between the two distributions.

where $z_k \sim \mathcal{N}(0, 1)$, N is the number of Monte Carlo samples, and the term $\log \pi(z_k)$ is omitted since its expectation is constant. The expectation of the data fitting loss and image regularization loss are optimized by sampling images from the generative model $G_\theta(\cdot)$. Note that when $\mathcal{L}(\cdot, \cdot)$ does not define the true data likelihood, or $\mathcal{R}(\cdot)$ does not define an image prior, the learned network only models an approximate image posterior instead of the true posterior distribution.

The data fitting loss and the regularization function are often empirically defined and may not match reality perfectly. Recalling that the third term is related to the entropy of the learned distribution, similar to β -VAE(Higgins et al. 2016), we introduce another tuning parameter β to control the diversity of the generative model samples,

$$\theta^* = \arg \min_{\theta} \sum_{k=1}^N \left\{ \mathcal{L}(y, f(G_\theta(z_k))) + \lambda \mathcal{R}(G_\theta(z_k)) - \beta \log \left| \det \frac{dG_\theta(z_k)}{dz_k} \right| \right\}. \quad (10)$$

When the uncertainty of the reconstructed images seems smaller than expected, we can increase β to encourage higher entropy of the generative distribution; otherwise, we can reduce β to reduce the diversity of reconstructions.

Larger β also encourages more exploration during training, which can be used to accelerate convergence.

Toy Examples

We first study our method using two-dimensional toy examples. Assuming x is two-dimensional, and the joint distribution $p(y, x)$ is given exactly by the potential function $\mathcal{J}(x)$, Eq. 10 can be simplified to

$$\theta^* = \arg \min_{\theta} \left\{ \mathbb{E}_{z \sim \mathcal{N}(0,1)} [\mathcal{J}(G_{\theta}(z))] - \beta \log \left| \det \frac{dG_{\theta}(z_k)}{dz_k} \right| \right\}. \quad (11)$$

For the following toy tests the generative model $G_{\theta}(\cdot)$ is designed using a Real-NVP architecture with 32 affine coupling layers (Dinh, Sohl-Dickstein, and Bengio 2016).

We test the approach on 3 joint distribution functions: (a) a Gaussian mixture potential, (b) a bowtie potential, and (c) a Sinusoidal potential. Fig. 1 shows the true and learned probability density function in these three cases for $\beta = 1$. Qualitatively, the learned generative model distributions match the true distributions well.

As derived in Eq. 7, the posterior should be best learned by $G_{\theta}(\cdot)$ for entropy weight $\beta = 1$. To test this claim, we adjust β from 0.1 to 10 to see how the entropy term influences the learned generative model distribution. According to the graph of KL divergence versus β in Fig. 1, the learned distributions match the true distributions best when the entropy weight equals 1. This verifies the theoretical expectation presented in the Method section.

Since the generative model is a transformation of a continuous multivariate Gaussian distribution, the learned distribution is also continuous. This leads to a common issue in flow-based generative models: there are often a few samples located in the high loss regions when the modes are not connected (see Fig. 1 distribution (a) and (b)). Some approaches (Gao et al. 2020) have been proposed recently to solve this problem, however, in this paper we neglect this issue and leave it for future work.

Interferometric Imaging Case Study

We demonstrate our proposed approach on the problem of radio interferometric astronomical imaging (Richard Thompson, Moran, and Swenson Jr 2017). In interferometric imaging, radio telescopes are linked to obtain sparse spatial frequency measurements of an astronomical target. These Fourier measurements are then used to recover the underlying astronomical image. Since the measurements are often very sparse and noisy, there can be multiple image modes that fit the observed data well. The Event Horizon Telescope (EHT) used this technique to take the first picture of a black hole, by linking telescopes across the globe (The EHT Collaboration et al. 2019). Fig. 2 shows the spatial frequency (Fourier) samples that can be measured by a 9-telescope EHT array when observing the black hole Sagittarius A* (Sgr A*)¹.

¹The largest sampled spatial frequency determines the interferometer’s nominal resolution of $\approx 25 \mu\text{as}$ for the EHT. In this paper, we neglect evolution of Sgr A* and assume it is static over the course of a night.

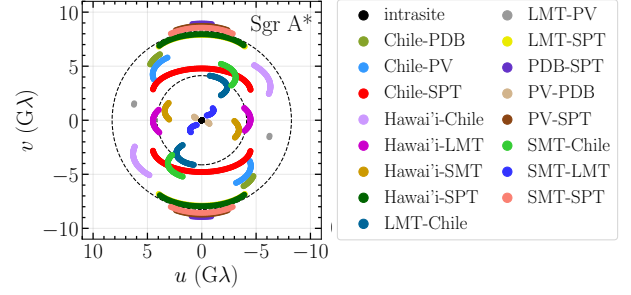


Figure 2: Spatial frequency samples for a 9-telescope EHT array observing the black hole Sgr A* over a night. The (u, v) position indicates the 2D spatial Fourier component of the image that is sampled by a telescope pair.

Each Fourier measurement, or so-called “visibility”, is obtained by correlating the data from a pair of telescopes. The measurement equation for each visibility is given by

$$V_{a,b}^t = g_a^t g_b^t \exp[-i(\phi_a^t - \phi_b^t)] F_{a,b}^t x + n_{a,b}^t, \quad (12)$$

where a and b index the telescopes, t represents time, and $F_{a,b}^t x$ extracts the Fourier component from image x corresponding to the baseline between telescope a and b at time t . The measurement noise comes from three sources: (1) time-dependent telescope-based gain error, g_a^t and g_b^t , (2) time-dependent telescope-based phase error, ϕ_a^t and ϕ_b^t , and (3) baseline-based Gaussian thermal noise, $n_{a,b}^t \sim \mathcal{N}(0, \nu_{a,b}^2)$. The standard derivation of thermal noise depends on each telescope’s System Equivalent Flux Density (SEFD):

$$\nu_{a,b} \propto \sqrt{SEFD_a \times SEFD_b}. \quad (13)$$

When the gain and phase errors are reasonably small, the interferometric imaging problem is approximately a convex inverse problem. However, when the gain errors and the phase errors (caused by atmospheric turbulence and instrument miscalibration) are large, the noisy visibilities can be combined into robust data products that are invariant to telescope-based errors, termed closure phase, $C_{a,b,c}^{\text{ph},t}$, and closure amplitude, $C_{a,b,c,d}^{\text{amp},t}$ (Chael et al. 2018):

$$\begin{aligned} C_{a,b,c}^{\text{ph},t} &= \angle (V_{a,b}^t V_{b,c}^t V_{c,a}^t) \approx f_{a,b,c}^{\text{ph},t}(x), \\ C_{a,b,c,d}^{\text{amp},t} &= \frac{|V_{a,b}^t| |V_{c,d}^t|}{|V_{a,c}^t| |V_{b,d}^t|} \approx f_{a,b,c,d}^{\text{amp},t}(x). \end{aligned} \quad (14)$$

These nonlinear “closure quantities” can be used to constrain non-convex image reconstruction

$$\begin{aligned} \mathcal{L}(y, f(x)) &= \sum_{a,b,c} |C_{a,b,c}^{\text{ph},t} - f_{a,b,c}^{\text{ph},t}(x)|^2 / \sigma_{a,b,c}^2 \\ &+ \sum_{a,b,c,d} |\log C_{a,b,c,d}^{\text{amp},t} - \log f_{a,b,c,d}^{\text{amp},t}(x)|^2 / \sigma_{a,b,c,d}^2, \end{aligned} \quad (15)$$

where $\sigma_{a,b,c}$ and $\sigma_{a,b,c,d}$ are the standard deviations of the corresponding closure term computed based on SEFDs. Note the closure quantities are not independent and the corresponding standard deviations are derived from linearization, so Eq. 15 only approximates the true data log-likelihood.

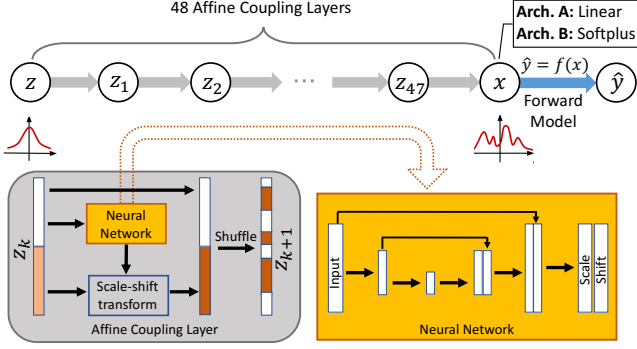


Figure 3: The Real-NVP network architecture used in DPI interferometric imaging. It consists of 48 affine coupling layers, transforming a latent Gaussian sample z to an image reconstruction sample, x . In each affine coupling layer, the input vector, z_k , is split into two parts: the first half is kept unchanged, while the second half is modified based on a neural network transformation of the first half. The network is composed of dense layers (including the Leaky ReLU activation and the batch normalization) with skip connections, similar to a UNet. After each affine-coupling transform block, the vector is randomly shuffled, so that different elements are modified in the next transform block.

In the following contents, we demonstrate our Deep Probabilistic Imaging (DPI) approach on both convex reconstruction with complex visibilities and non-convex reconstruction with closure quantities using both synthetic and real datasets. With this new approach, we successfully quantify the uncertainty of interferometric images, as well as detect multiple modes in some data sets.

Implementation For all interferometric imaging DPI results we use a Real-NVP (Dinh, Sohl-Dickstein, and Bengio 2016) network architecture with 48 affine coupling layers (Fig. 3). The scale-shift function of each layer is modeled as a U-Net style fully connected network with five hidden layers and skip connections. We train the model using Adam with a batch size of 32 for 20,000 epochs. We note that a limitation of our general approach is the modeling capacity of the flow-based generative model used. We find the proposed architecture is satisfactory for characterizing uncertainty in images of size 32×32 pixels.

Convex Imaging with Visibilities

In this section, we demonstrate DPI on convex interferometric imaging. In particular, the gain and phase errors in Eq. 12 are assumed to be zero (i.e., $g_a^t = g_b^t = 1$ and $\phi_a^t = \phi_b^t = 0$) so that complex visibilities are a linear function of the latent image. Since the thermal noise on $V_{a,b}^t$ is independent and Gaussian, we write the conditional likelihood as

$$\mathcal{L}(y, f(x)) = \frac{1}{2}(y - Fx)^T \Sigma^{-1}(y - Fx), \quad (16)$$

where x is a column vectorized image with M^2 pixels, $y = [\dots, \Re\{V_{a,b}^t\}, \Im\{V_{a,b}^t\}, \dots]$ is a column vector of K complex visibility measurements, $F = [\dots, \Re\{F_{a,b}^t\}, \Im\{F_{a,b}^t\}, \dots]$ is a under-sampled Fourier transform matrix of size $K \times M^2$, and $\Sigma =$

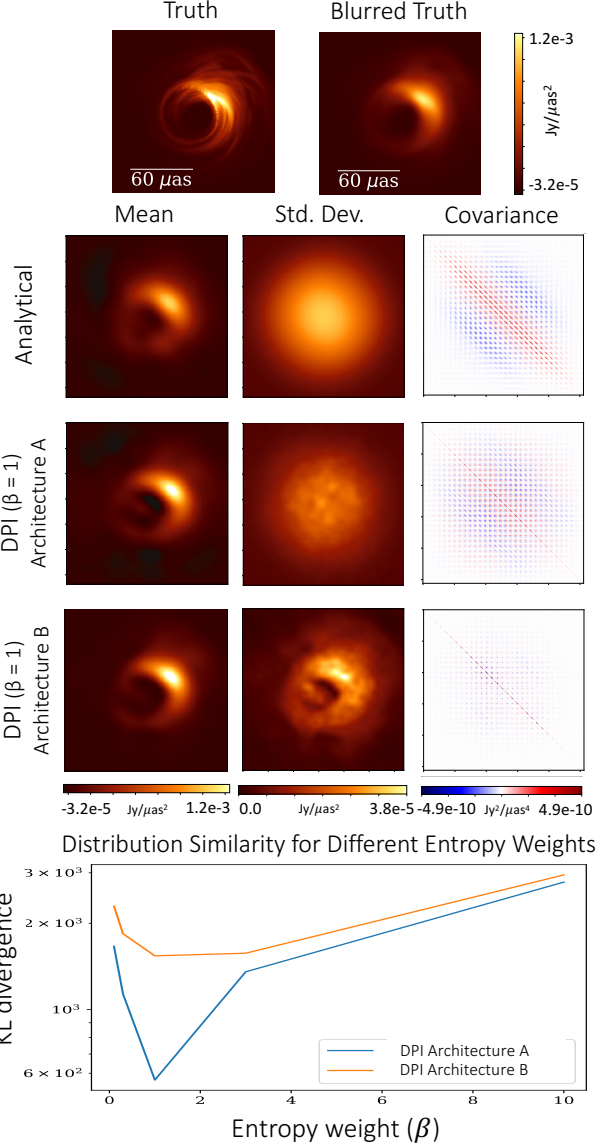


Figure 4: Posterior distribution estimation for convex interferometric image reconstruction. (Top) The target synthetic image of a black hole (original and blurred to $0.4 \times$ the interferometer's nominal resolution of $\approx 25 \mu\text{as}$). Noisy visibility measurements are derived from this target image. (Middle) The posterior mean, pixel-wise standard deviation, and full covariance obtained analytically using a Gaussian image prior and with two DPI generative model architectures. Architecture A allows negative pixel values, while Architecture B restricts images to be non-negative. The DPI results are computed according to 2048 samples from each generative model. (Bottom) The similarity between the DPI distributions and the analytical distribution under different entropy weights β . Both architectures achieve a minimum at $\beta = 1$.

$\text{diag}([\dots, \nu_{a,b}^2, \nu_{a,b}^2, \dots])$ is a $K \times K$ measurement covariance matrix derived according to the telescopes' SEFD.

In order to verify that the flow-based generative model can learn the posterior distribution of reconstructed images, we

employ a multivariate Gaussian image prior,

$$\mathcal{R}(x) = \frac{1}{2}(x - \mu)^T \Lambda^{-1}(x - \mu), \quad (17)$$

where μ is a mean image, and Λ is the covariance matrix defined by the empirical power spectrum of an image (Bouman et al. 2018). Since both the measurement and image prior follow Gaussian distributions, the reconstructed image’s posterior distribution is also a Gaussian distribution and can be analytically derived as

$$\begin{aligned} p(x|y) &\propto p(y|x)p(x) = \mathcal{N}_y(Fx, \Sigma)\mathcal{N}_x(\mu, \Lambda) = \mathcal{N}_x(m, C) \\ m &= \mu + \Lambda F^T(\Sigma + F\Lambda F^T)^{-1}(y - F\mu) \\ C &= \Lambda - \Lambda F^T(\Sigma + F\Lambda F^T)^{-1}F\Lambda. \end{aligned} \quad (18)$$

Using the specified data likelihood and prior, we train a DPI flow-based network to produce image samples of size 32×32 pixels with a field of view of 160 micro-arcseconds (μas). Fig. 4 demonstrates DPI on a synthetic interferometric imaging example, and compares the learned generative model distribution with the analytical posterior distribution. Visibility measurements are derived from the synthetic black hole image shown in the top of Fig. 4 (with a total flux of 2 Janskys). The second row of the Fig. 4 shows the analytic posterior’s mean, standard deviation, and full covariance. The third and fourth rows of the figure show the mean, standard deviation, and full covariance empirically derived from DPI samples under two slightly different $G_\theta(\cdot)$ architectures: (A) the third row uses a model with 48 affine coupling layers, and (B) the fourth row adds an additional Soft-plus layer to the model to enforce the physical constraint of non-negativity in the image distribution. Without a non-negativity constraint, Architecture A’s learned distribution is very similar to the analytical posterior distribution, since it is solving a same Bayesian estimation problem as defined in Eq. 18. However, this Bayesian estimation problem does not constrain the image to be positive; as a result, the central depression in the image has an inflated standard deviation. Architecture B’s non-negative model results in a more constraining uncertainty map while achieving a slightly higher resolution reconstruction. This example also demonstrates how DPI can introduce implicit regularization through the neural network architecture.

Image distributions with different levels of sample diversity can be learned by adjusting the entropy loss weight, β . As expected, both generative models reach lowest KL divergence with the analytic distribution when $\beta = 1$.

Non-convex Imaging with Closure Quantities

In this section, we demonstrate DPI on non-convex interferometric imaging, where we reconstruct images using the closure quantities defined in Eq. 14. With this non-convex forward model, the posterior of reconstructed images cannot be analytically computed, but it can be estimated using DPI. In all DPI reconstructions, the resulting images are 32×32 pixels and result from the non-negative Real-NVP model discussed above (Architecture B).

Multi-modal Posterior Distributions A serious challenge for non-convex image reconstruction is the potential

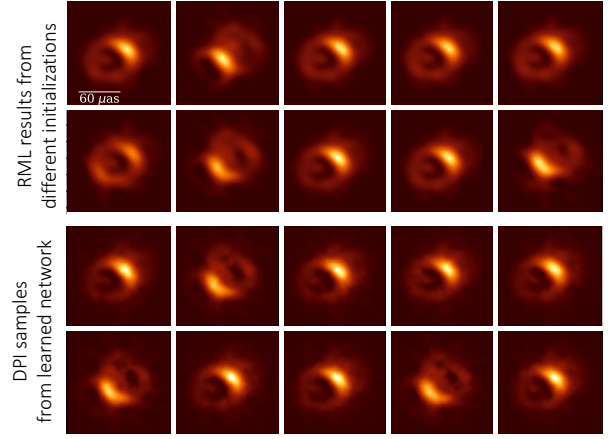


Figure 5: Non-convex interferometric imaging results with closure quantities. (Top) RML reconstructed images obtained from different initializations. (Bottom) Samples from a learned DPI flow-based generative model.

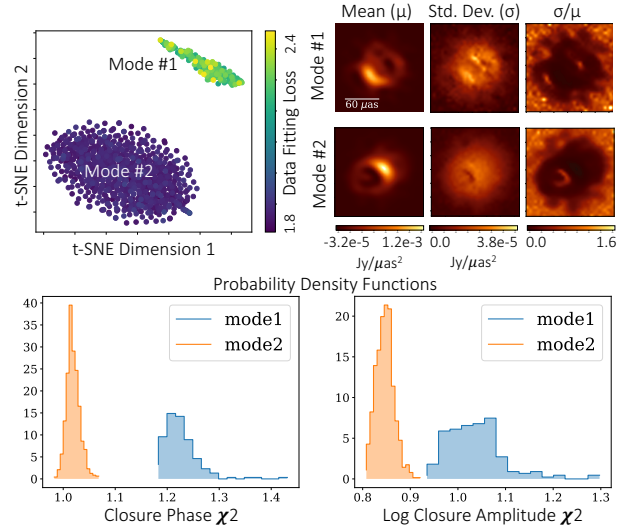


Figure 6: Analysis of 1024 reconstructed images sampled from a DPI generative model trained with closure quantities. (Top Left) two-dimensional t-SNE plot of samples with perplexity= 20. The samples clearly cluster into two modes. (Top Right) The mean, standard deviation, and fractional standard deviation for samples from each mode. (Bottom) The distributions of data fitting losses (reduced χ^2) of samples from each mode. A χ^2 value of 1 is optimal for high SNR data. The second mode, which happens to be the correct solution, results in a distribution with smaller data fitting losses.

for multi-modal posterior distributions: visually-different solutions fit the measurement data reasonably well. In some cases, multiple modes can be identified by optimizing a regularized maximum likelihood (RML) loss with different image initializations; for example, Fig. 5 (top) shows ten RML reconstructed images obtained using the closure quantities (Eq. 15) from the target shown in Fig. 4 and the multivariate Gaussian regularizer defined in Eq. 17. From these results two potential image modes, which appear to be roughly 180 degree rotations of one another, clearly stand out as fitting

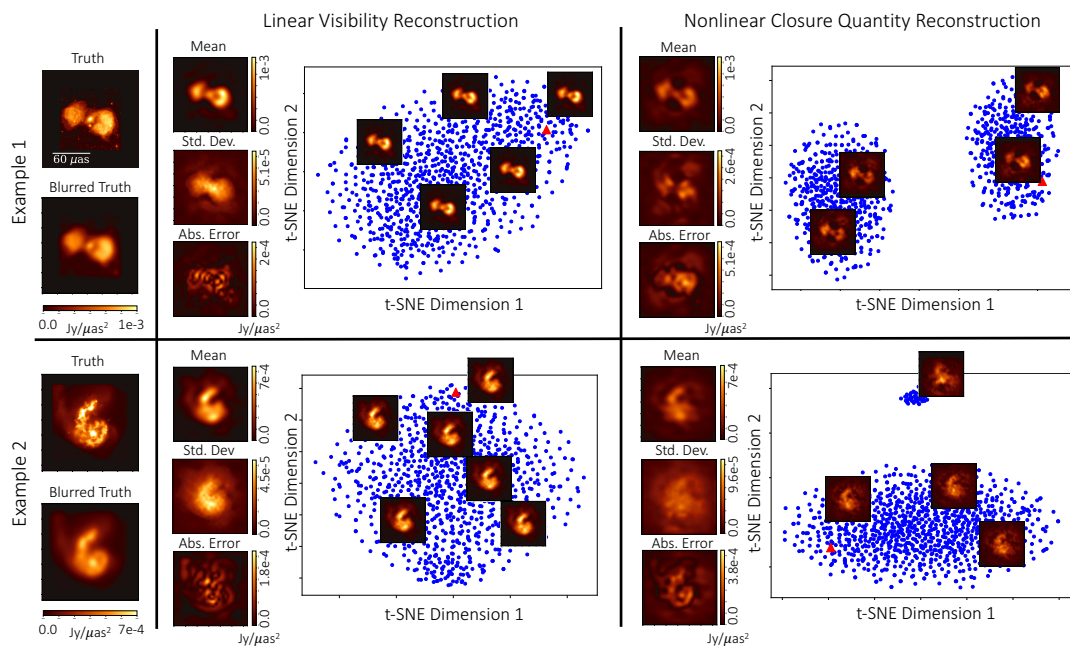


Figure 7: Examples of uncertainty quantification in convex and non-convex interferometric imaging with multivariate Gaussian regularization. In each case, the mean, standard deviation, and absolute error between the reconstruction mean and blurred truth are reported. t-SNE plots (perplexity= 20) are used to visualize the distributions of 1024 image samples in a two-dimensional embedded space. Within the t-SNE plots, each small image corresponds to a sample embedded at its bottom-left corner. In convex interferometric imaging the area of high error approximately matches the uncertainty estimated by DPI. In non-convex interferometric imaging, both examples produced two solution modes. The red triangle marks the embedding of the blurred truth image, which appears close to samples in the embedded image space.

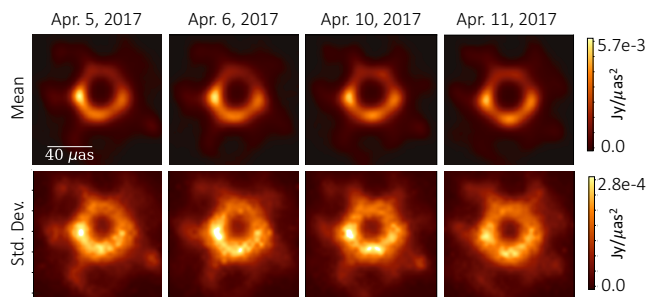


Figure 8: DPI imaging results using real EHT 2017 observation data of the black hole in M87 on different days. See (The EHT Collaboration et al. 2019) for the (u, v) spatial frequency coverage for each day.

the data well.

Fig. 5 (bottom) shows ten images sampled from a DPI flow-based generative model learned with multivariate Gaussian regularization. Note that the single generative model has captured the two modes identified by multiple runs of RML.

Fig. 6 analyzes 1024 generative samples from a DPI model learned with multivariate Gaussian regularization. The dimensionality reduction t-SNE plot (Maaten and Hinton 2008) indicates a clustering of samples into two modes. Figure 6 (top right) shows the pixel-wise mean, standard deviation and fractional standard deviation of samples for each mode. The distributions of data fitting loss (reduced χ^2) for images in each mode are shown for both closure phase and

log closure amplitude constraints; a reduced χ^2 value of 1 is optimal. Although it can be difficult to tell which image is correct by inspecting the statistics of a single image, by analyzing the histogram of statistics for each mode it becomes clearer which mode corresponds with the true image. In the supplemental material² we show how the resulting posterior changes under different imposed regularization.

Real Interferometric Data In Fig. 8 we demonstrate the performance of DPI using the publicly available EHT 2017 data, which was used to produce the first image of the black hole in M87. In accordance with (The EHT Collaboration et al. 2019), we use a data fitting loss with not only closure quantities (Eq. 15) but also roughly-calibrated visibility amplitudes. We pair this data likelihood with a combined maximum entropy (MEM) and total squared variation (TSV) regularizer (see the supplementary material for details²). Fig. 8 shows the DPI reconstruction mean and standard deviation of M87 on different observation days. Although ground truth images are unavailable, the ring size and evolution recovered using DPI matches that seen in the original EHT imaging results. The DPI results also quantify larger uncertainty in “knot” regions along the lower portion of the ring.

Visualizing Uncertainty

DPI sampling provides a means to visualizing uncertainty, especially in cases where closed form approximations are insufficient. By embedding samples from our DPI model in a two-dimensional space, we are able to visualize the pos-

² <http://imaging.cms.caltech.edu/dpi/DPIsupplement.pdf>

terior’s approximate structure. Fig. 7 plots the embedding of DPI samples obtained using t-SNE, and shows the images corresponding to some of these samples. By plotting the blurred truth image in the same embedding we see that the truth often lies close to the posterior samples in the embedded space, though it is not guaranteed. The posterior is sensitive to the choice of image prior, and is most informative when the true image is a sample from the imposed prior.

The mean and standard deviation of DPI samples, as well as the average absolute error with respect to the blurred truth, are shown for each learned distribution in Fig. 7. Since closure quantities do not constrain the absolute flux or position of the reconstructed source, we first normalize and align samples from the closure-constrained DPI model to account for scaled and shifted copies of similar images. Note that the pixel-wise standard deviation aligns well with areas of high error in the generative model’s samples.

Compressed Sensing MRI Case Study

Under-sampled measurements in compressed sensing MRI also result in image reconstruction uncertainty. Similar to convex interferometric imaging with visibilities, compressed sensing MRI is often modeled as a linear forward model, $y = Fx + \epsilon$, where the measurements y are the under-sampled κ -space spatial frequency components of the image x with additive noise ϵ . In this section, we apply DPI to compressed sensing MRI data with different acceleration speed-up factors and compare the DPI identified uncertainty map to the image reconstruction errors.

Figure 9 shows the pixel-wise statistics of the DPI estimated posterior (computed from 1024 generated images) of a synthetic example (a knee image from fastMRI dataset (Zbontar et al. 2018) resized to 64×64 pixels). The κ -space measurement noise is assumed Gaussian with a standard deviation of 0.04% of the DC (zero-frequency) amplitude. DPI is tested at three different acceleration speed-up factors: $3.5\times$, $5.5\times$ and $8.4\times$, i.e. $1/3.5$, $1/5.5$ or $1/8.4$ of all κ -space components are observed. A Real-NVP network architecture similar to Fig. 3 with 32 affine coupling layers is used to approximate the image posterior, and a total variation (TV) regularizer is applied as the image prior. As expected, the pixel-wise standard deviation of DPI reconstruction becomes larger as the acceleration speed increases. The reconstruction values of most pixels lie within four standard deviations (σ) from the truth (98.2% pixels for $3.5\times$, 98.6% pixels for $5.5\times$, and 98.7% pixels for $8.4\times$ are within 4σ). Since the image posterior distribution is not pixel-wise independent Gaussian, the learned reconstruction error should not necessarily obey properties of Gaussian statistics. The profile of DPI standard deviation estimation roughly captures the pattern of the absolute reconstruction error in all cases.

Conclusion

In this paper, we present deep probabilistic imaging (DPI): a new framework for uncertainty quantification and multi-modal solution characterization for underdetermined image reconstruction problems. The method parameterizes the posterior distribution of the reconstructed image as an untrained flow-based generative model, and learns the neural network

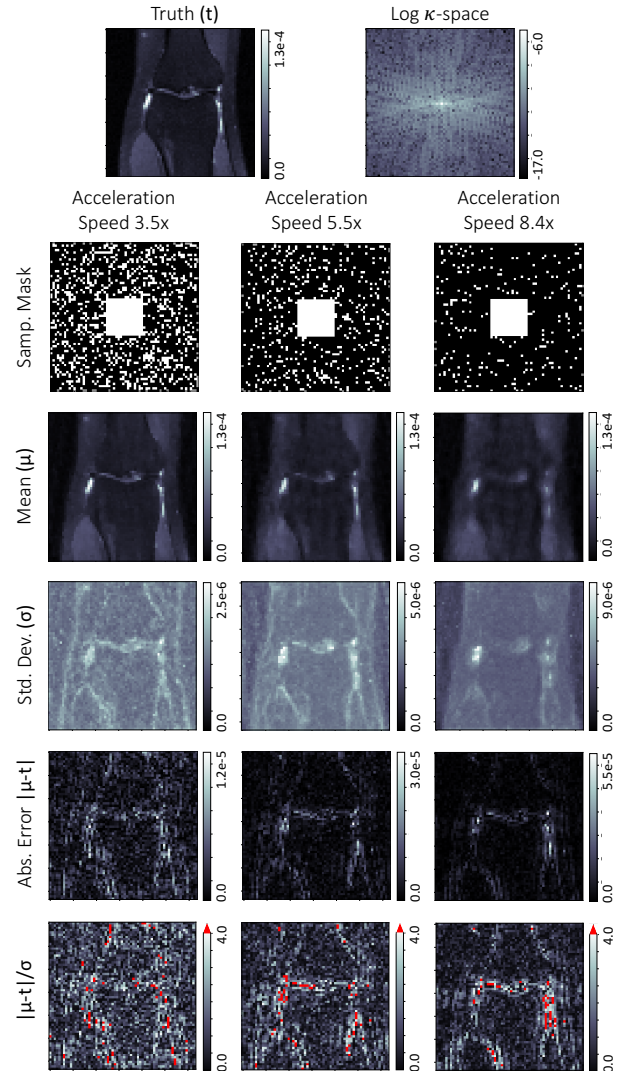


Figure 9: DPI imaging results of a synthetic compressed sensing MRI example. A knee image from the fastMRI dataset (Zbontar et al. 2018) is tested at three acceleration speed-up factors: $3.5\times$, $5.5\times$ and $8.4\times$ (each shown in a column). White pixels in the sampling masks (second row) indicate the observed κ -space components, where the DC (zero-frequency) component is aligned with the center of the mask. According to the pixel-wise statistics of the estimated posterior distributions (rows three and four), DPI well identifies the highly uncertain areas in the reconstructed images.

weights using a loss that incorporates the conditional data likelihood, prior of image samples, and the model’s distribution entropy.

We demonstrate the proposed method on toy examples, synthetic and real interferometric imaging problems, as well as a synthetic compressed sensing MRI problem. Experiments show the proposed method can approximately learn the image posterior distribution in both convex and non-convex inverse problems, which enables efficiently quantifying the uncertainty of reconstructed images and detecting multi-modal solutions. Code is available at <http://imaging.cms.caltech.edu/dpi/>.

Acknowledgments

This work was supported by NSF award 1935980: Next Generation Event Horizon Telescope Design, and Beyond Limits. The authors would also like to thank Joe Marino, Dominic Pesce, S. Kevin Zhou, and Tianwei Yin for the helpful discussions.

References

- Arras, P.; Frank, P.; Leike, R.; Westermann, R.; and Enßlin, T. A. 2019. Unified radio interferometric calibration and imaging with joint uncertainty quantification. *A&A* 627: A134. doi:10.1051/0004-6361/201935555.
- Bardsley, J. M. 2012. MCMC-based image reconstruction with uncertainty quantification. *SIAM Journal on Scientific Computing* 34(3): A1316–A1332.
- Blei, D. M.; Kucukelbir, A.; and McAuliffe, J. D. 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association* 112(518): 859–877.
- Bostan, E.; Heckel, R.; Chen, M.; Kellman, M.; and Waller, L. 2020. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. *Optica* 7(6): 559–562.
- Bouman, C.; and Sauer, K. 1993. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on image processing* 2(3): 296–310.
- Bouman, K. L.; Johnson, M. D.; Dalca, A. V.; Chael, A. A.; Roelofs, F.; Doeleman, S. S.; and Freeman, W. T. 2018. Reconstructing video of time-varying sources from radio interferometric measurements. *IEEE Transactions on Computational Imaging* 4(4): 512–527.
- Broderick, A. E.; Pesce, D. W.; Tiede, P.; Pu, H.-Y.; and Gold, R. 2020. Hybrid Very Long Baseline Interferometry Imaging and Modeling with themis. *The Astrophysical Journal* 898(1): 9.
- Cai, X.; Pereyra, M.; and McEwen, J. D. 2018. Uncertainty quantification for radio interferometric imaging—I. Proximal MCMC methods. *Monthly Notices of the Royal Astronomical Society* 480(3): 4154–4169.
- Candes, E.; and Romberg, J. 2007. Sparsity and incoherence in compressive sampling. *Inverse problems* 23(3): 969.
- Chael, A. A.; Johnson, M. D.; Bouman, K. L.; Blackburn, L. L.; Akiyama, K.; and Narayan, R. 2018. Interferometric imaging directly with closure phases and closure amplitudes. *The Astrophysical Journal* 857(1): 23.
- Dinh, L.; Krueger, D.; and Bengio, Y. 2014. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*.
- Dinh, L.; Sohl-Dickstein, J.; and Bengio, S. 2016. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*.
- Gal, Y. 2016. Uncertainty in deep learning. *University of Cambridge* 1(3).
- Gao, R.; Nijkamp, E.; Kingma, D. P.; Xu, Z.; Dai, A. M.; and Wu, Y. N. 2020. Flow contrastive estimation of energy-based models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7518–7528.
- Gershman, S.; Hoffman, M.; and Blei, D. 2012. Nonparametric variational inference. *arXiv preprint arXiv:1206.4665*.
- Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S.; and Lerchner, A. 2016. beta-vae: Learning basic visual concepts with a constrained variational framework.
- Jordan, M. I.; Ghahramani, Z.; Jaakkola, T. S.; and Saul, L. K. 1999. An introduction to variational methods for graphical models. *Machine learning* 37(2): 183–233.
- Kingma, D. P.; and Dhariwal, P. 2018. Glow: Generative flow with invertible 1x1 convolutions. In *Advances in neural information processing systems*, 10215–10224.
- Kuramochi, K.; Akiyama, K.; Ikeda, S.; Tazaki, F.; Fish, V. L.; Pu, H.-Y.; Asada, K.; and Honma, M. 2018. Superresolution interferometric imaging with sparse modeling using total squared variation: application to imaging the black hole shadow. *The Astrophysical Journal* 858(1): 56.
- Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9(Nov): 2579–2605.
- MacKay, D. J. 1995. Bayesian neural networks and density networks. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 354(1): 73–80.
- Ongie, G.; Jalal, A.; Baraniuk, C. A. M. R. G.; Dimakis, A. G.; and Willett, R. 2020. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*.
- Repetti, A.; Pereyra, M.; and Wiaux, Y. 2019. Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization. *SIAM Journal on Imaging Sciences* 12(1): 87–118.
- Rezende, D. J.; and Mohamed, S. 2015. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*.
- Richard Thompson, A.; Moran, J. M.; and Swenson Jr, G. W. 2017. *Interferometry and synthesis in radio astronomy*. Springer Nature.
- Skilling, J.; and Bryan, R. 1984. Maximum entropy image reconstruction-general algorithm. *Monthly notices of the royal astronomical society* 211: 111.
- The EHT Collaboration, Akiyama, K.; Alberdi, A.; Alef, W.; Asada, K.; Azulay, R.; Baczko, A.-K.; Ball, D.; Baloković, M.; Barrett, J.; Bintley, D.; et al. 2019. First M87 event horizon telescope results. IV. Imaging the central supermassive black hole. *The Astrophysical Journal Letters* 875(1): L4.
- Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2018. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9446–9454.

Xue, Y.; Cheng, S.; Li, Y.; and Tian, L. 2019. Reliable deep-learning-based phase imaging with uncertainty quantification. *Optica* 6(5): 618–629.

Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M. J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.

Zoran, D.; and Weiss, Y. 2011. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, 479–486. IEEE.