Discovering Structure From Corruption for Unsupervised Image Reconstruction

Oscar Leong [ID], *Member, IEEE*, Angela F. Gao [ID], He Sun [ID], *Member, IEEE*, and Katherine L. Bouman [ID]

Abstract—We consider solving ill-posed imaging inverse problems without access to an image prior or ground-truth examples. An overarching challenge in these inverse problems is that an infinite number of images, including many that are implausible, are consistent with the observed measurements. Thus, image priors are required to reduce the space of possible solutions to more desirable reconstructions. However, in many applications it is difficult or potentially impossible to obtain example images to construct an image prior. Hence inaccurate priors are often used, which inevitably result in biased solutions. Rather than solving an inverse problem using priors that encode the spatial structure of any one image, we propose to solve a set of inverse problems jointly by incorporating prior constraints on the collective structure of the underlying images. The key assumption of our work is that the underlying images we aim to reconstruct share common, low-dimensional structure. We show that such a set of inverse problems can be solved simultaneously without the use of a spatial image prior by instead inferring a shared image generator with a low-dimensional latent space. The parameters of the generator and latent embeddings are found by maximizing a proxy for the Evidence Lower Bound (ELBO). Once identified, the generator and latent embeddings can be combined to provide reconstructed images for each inverse problem. The framework we propose can handle general forward model corruptions, and we show that measurements derived from only a small number of ground-truth images ($\leqslant 150$) are sufficient for image reconstruction. We demonstrate our approach on a variety of convex and non-convex inverse problems, including denoising, phase retrieval, and black hole video reconstruction.

Index Terms—Inverse problems, computational imaging, prior models, generative networks, Bayesian inference.

I. INTRODUCTION

IN IMAGING inverse problems, the goal is to recover an underlying image from corrupted measurements, where the measurements and image are related via an understood forward model: $y = f(x) + \eta$. Here, y are measurements, x is the underlying image, f is a known forward model, and η is noise. Such problems are ubiquitous and include denoising [6], [20], super-resolution [7], compressed sensing [8], [19], phase retrieval [21], and deconvolution [35]. Due to corruption by the forward model and noise, these problems are often ill-posed: there are many images that are consistent with the observed measurements, including ones that are implausible.

To combat the ill-posedness in imaging problems, solving for an image traditionally requires imposing additional structural assumptions to reduce the space of possible solutions. We encode these assumptions in an *image generation model (IGM)*, whose goal is to capture the desired properties of an image's spatial structure. IGMs are general; they encompass probabilistic *spatial-domain* priors (e.g., that encourage smoothness or sparsity), but also include deep image generators that are not necessarily probabilistic but are trained to primarily sample a certain class of images.

In order to define an IGM, it is necessary to have knowledge of the underlying image's structure. If images similar to the underlying image are available, then an IGM can be learned directly [4], [44], [56]. However, an abundance of clean images is not available for many scientific imaging modalities (e.g., geophysical imaging and astronomical imaging). Collecting images in these domains can be extremely invasive, time-consuming, expensive, or even impossible. For instance, how should we define an IGM for black hole imaging without having ever seen a direct image of a black hole or knowing what one should look like? Moreover, classical approaches that utilize hand-crafted IGMs, such as total variation [26] or sparsity in a wavelet basis [37], are prone to human bias [34].

In this work, we show how one can solve a set of ill-posed image reconstruction tasks in an unsupervised fashion, i.e., without prior information about an image's spatial structure or access to clean, example images. The key insight of our work is that knowledge of common structure across multiple diverse images can be sufficient regularization alone. In particular, suppose we have access to a collection of noisy measurements $\{y^{(i)}\}_{i=1}^N$ that are observed through (potentially different) forward models $y^{(i)} := f^{(i)}(x^{(i)}) + \eta^{(i)}$. The core assumption we make is that the different underlying images $\{x^{(i)}\}_{i=1}^N$ are drawn from the same distribution (unknown *a priori*) and share

common, low-dimensional structure. Thus, our "prior" is not at the spatial-level, but rather exploits the *collective structure* of the underlying images. This assumption is satisfied in a number of applications where there is no access to an abundance of clean images. For instance, although we might not know what a black hole looks like, we might expect it to be similar in appearance over time. We show that under this assumption, the image reconstruction posteriors $p(x|y^{(i)})$ can be learned jointly from a small number of examples $\{y^{(i)}\}_{i=1}^{N}$ due to the common, low-dimensional structure of the collection $\{x^{(i)}\}_{i=1}^{N}$. Specifically, our main result is that one can capitalize on this common structure by jointly solving for 1) a shared image generator G_θ and 2) N low-dimensional latent distributions $q_{\phi^{(i)}}$, such that the distribution induced by the push-forward of $q_{\phi^{(i)}}$ through G_θ approximately captures the image reconstruction posterior $p(x|y^{(i)})$ for each measurement example $i \in [N]$.

A. Our Contributions

We outline the main contributions of our work, which extends our prior work presented in [22]:
1) We solve a collection of ill-posed inverse problems without prior knowledge of an image's spatial structure by exploiting the common, low-dimensional structure shared across images. This common structure is exploited when inferring a shared IGM with a low-dimensional latent space.
2) To infer this IGM, we define a loss inspired by the evidence lower bound (ELBO). We motivate this loss by showing how it aids in unsupervised image reconstruction by helping select one IGM from a collection of candidate IGMs using a single measurement example.
3) We apply our approach to convex and non-convex inverse problems, such as denoising, black hole compressed sensing, and phase retrieval. We establish that we can solve inverse problems without spatial-level priors and demonstrate good performance with only a small number of independent measurement examples (e.g., $\leqslant 150$).
4) We theoretically analyze the inferred IGM in linear inverse problems under a linear image model to show that in this setting the inferred IGM performs dimensionality reduction akin to PCA on the collection of measurements.

II. BACKGROUND AND RELATED WORK

We now discuss related literature in model selection and learning-based IGMs. In order to highlight our key contributions, we emphasize the following assumptions in our framework:
1) We do not have access to a set of images from the same distribution as the underlying images.
2) We only have access to a collection of measurement examples, where each example comes from a different underlying image. The number of examples N is small, e.g., $N \leqslant 150$.
3) For each underlying image $x^{(i)}$ we wish to reconstruct, we only have access to a single measurement example $y^{(i)} = f^{(i)}(x^{(i)}) + \eta^{(i)}$. That is, we do not have multiple

observations of the same underlying image. Note each $f^{(i)}$ can be potentially different.

A. Model Selection

Model selection techniques seek to choose a model that best explains data by balancing performance and model complexity. In supervised learning problems with sufficiently large amounts of data, this can be achieved simply by evaluating the performance of different candidate models using reserved test data [50]. However, in image reconstruction or other inverse problems with limited data, one cannot afford to hold out data. In these cases, model selection is commonly conducted using probabilistic metrics. The simplest probabilistic metric used for linear model selection is adjusted R^2 [40]. It re-weights the goodness-of-fit by the number of linear model parameters, helping reject high-dimensional parameters that do not improve the data fitting accuracy. Similar metrics in nonlinear model selection are Bayesian Information Criterion (BIC) [46] and Akaike Information Criterion (AIC) [1]. AIC and BIC compute different weighted summations of a model's log-likelihood and complexity, offering different trade-offs between bias and variance to identify the best model for a given dataset.

In our work, we consider the use of the ELBO as a model selection criterion. In [10], [11], the use of the ELBO as a model selection criterion is theoretically analyzed and rates of convergence for variational posterior estimation are shown. Additionally, [52] proposes a generalized class of evidence lower bounds leveraging an extension of the evidence score. In [51], the ELBO is used for model selection to select a few, discrete parameters modeling a physical system (e.g., parameters that govern the orbit of an exoplanet). A significant difference in our context, however, is that we use the ELBO as a model selection criterion in a *high-dimensional* imaging context, and we optimize the ELBO over a continuous space of possible parameters.

B. Learning IGMs

With access to a large corpus of example images, it is possible to directly learn an IGM to help solve inverse problems. Seminal work along these lines utilizing generative networks showcased that a pre-trained Generative Adversarial Network (GAN) can be used as an IGM in the problem of compressed sensing [4]. To solve the inverse problem, the GAN was used to constrain the search space for inversion. This approach was shown to outperform sparsity-based techniques with 5-10x fewer measurements. Since then, this idea has been expanded to other inverse problems, including denoising [25], super-resolution [39], magnetic resonance imaging (MRI) [2], [48], and phase retrieval [23], [47]. However, the biggest downside to this approach is the requirement of a large dataset of example images similar to the underlying image, which is often difficult or impossible to obtain in practice. Hence, we consider approaches that are able to directly solve inverse problems without example images.

Methods that aim to learn an IGM from only noisy measurements have been proposed. The main four distinctions between our work and these methods are that these works either: 1)

require multiple independent observations of the *same* underlying image, 2) can only be applied to certain inverse problems, 3) require significantly more observations (either through more observations of each underlying image or by observing more underlying images), or 4) require significant hyperparameter tuning based on knowledge of example images.

Noise2Noise (N2N) [33] learns to denoise images by training on a collection of noisy, independent observations of the *same* image. To do so, N2N learns a neural network Φ_θ whose goal is to map between noisy images y and denoised images x. Since it has no denoised image examples to supervise training, it instead employs a loss that maps between noisy examples of the same underlying image. This objective is as follows:

$$\arg\min_\theta \sum_{i=1}^{N} \mathbb{E}_{y_1^{(i)} \sim \mathcal{Y}_1} \mathbb{E}_{y_2^{(i)} \sim \mathcal{Y}_2} [L(\Phi_\theta(y_1^{(i)}), y_2^{(i)})], \quad (1)$$

where $y_j^{(i)}$ corresponds to a noisy observation of the i-th underlying image $x^{(i)}$, and $\mathcal{Y}_j$ is a distribution of noisy images where $\mathbb{E}_{y \sim \mathcal{Y}_j}[y] = x$. This N2N objective requires at least two observations of the same image and is limited by the assumption that the expected value of multiple observations of a single image is the underlying image. Thus, N2N is only applicable to denoising problems where the forward model is the identity matrix with independent noise on each pixel. Additionally, in practice N2N requires thousands of underlying images (i.e., $N = O(1000)$) to perform well. Thus, N2N's main distinctions with our work are distinctions 1), 2), and 3).

Regularization by Artifact Removal (RARE) [36] generalizes N2N to perform image reconstruction from measurements under linear forward models. That is, the objective in (1) is modified to include a pseudo-inverse. Nonetheless, multiple observations of the same underlying image are required, such that $\mathbb{E}_{y \sim \mathcal{Y}}[A^\dagger y] = x$ for the pseudo-inverse matrix $A^\dagger$. Thus, RARE suffers from the same limiting distinctions as N2N (i.e., 1), 2), and 3)).

Noise2Void [30] **and Noise2Self** [3] assume that the image can be partitioned such that the measurement noise in one subset of the partition is independent conditioned on the measurements in the other subset. This is true for denoising, but not applicable to general forward models. For example, in black hole and MRI compressed sensing, it is not true that the measurement noise can be independently partitioned since each measurement is a linear combination of all pixels. While this makes Noise2Void and Noise2Self more restrictive in the corruptions they can handle compared to RARE, they also don't require multiple observations of the same underlying image. Hence the main differences between these works and our own are distinctions 2) and 3).

AmbientGAN [5] and other similar approaches based on GANs [28] and Variational Autoencoders (VAEs) [38], [41] have been proposed to learn an IGM directly from noisy measurements. For instance, AmbientGAN aims to learn a generator whose images lead to simulated measurements that are indistinguishable from the observed measurements; this generator can subsequently be used as a prior to solve inverse problems. However, AmbientGAN requires many measurement examples

(on the order of 10,000) to produce a high quality generator. We corroborate this with experiments in Section IV to show that they require many independent observations and/or fine tuning of learning parameters to achieve good performance. Thus, the main distinctions between AmbientGAN and our work are 3) and 4).

Deep Image Prior (DIP) [55] uses a convolutional neural network as an implicit "prior". DIP has shown strong performance across a variety of inverse problems to perform image reconstruction without explicit probabilistic priors. However, it is prone to overfitting and requires selecting a specific stopping criterion. While this works well when example images exist, selecting this stopping condition from noisy measurements alone introduces significant human bias that can negatively impact results. Thus, the main distinction between DIP and our work is 4).

We would also like to highlight additional work done to improve certain aspects of the DIP. While the original DIP method used a U-Net architecture [45], other works such as the Deep Decoder [24] and ConvDecoder [17] used a decoder-like architecture that progressively grows a low-dimensional random tensor to a high-dimensional image. When underparameterized, such architectures have been shown to avoid overfitting and the need for early stopping. Other works mitigating early stopping include [9], which takes a Bayesian perspective to the DIP by using Langevin dynamics to perform posterior inference over the weights to improve performance and show this does not lead to overfitting.

III. APPROACH

In this work, we propose to solve a set of inverse problems without prior access to an IGM by assuming that the set of underlying images have common, low-dimensional structure. We motivate the use of optimizing the ELBO to infer an IGM by showing that it is a good criterion for generative model *selection* in Section III-A. Then, by optimizing the ELBO, we show in Section III-B that one can directly *infer* an IGM from corrupted measurements alone by parameterizing the image model as a deep generative network with a low-dimensional latent distribution. The IGM network weights are shared across all images, capitalizing on the common structure present in the data, while the parameters of each latent distribution are learned jointly with the generator to model the image posteriors for each measurement example. See Fig. 1 for a method overview.

A. Motivation for ELBO as a Model Selection Criterion

In order to accurately *infer* an IGM, we motivate the use of the ELBO as a loss by showing that it provides a principled criterion for *selecting* an IGM to use as a prior model. Suppose we are given noisy measurements from a single image: $y = f(x) + \eta$. In order to reconstruct the image x, we traditionally first require an IGM G that captures the distribution x was sampled from. A natural approach would be to find or select the model G that maximizes the model posterior distribution $p(G|y) \propto p(y|G)p(G)$. That is, conditioned on the noisy measurements, find the IGM
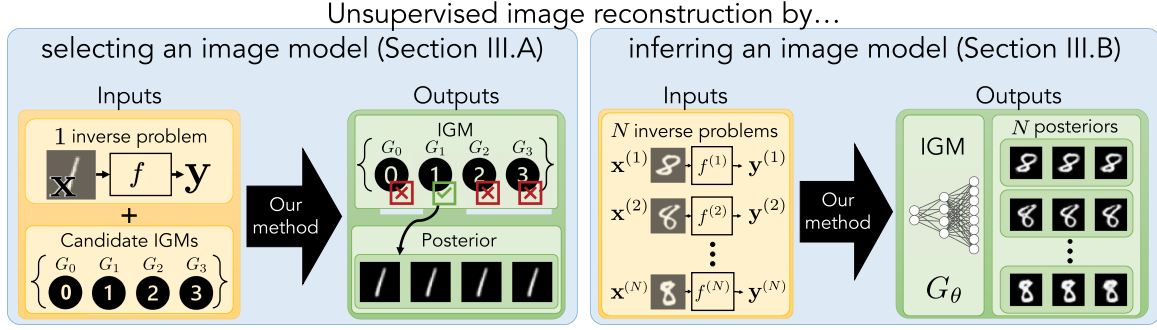
Fig. 1. Method overview. In this paper we tackle ill-posed image reconstruction problems when a traditional image prior is not available or cannot be derived from example images. The key idea of our work is that image reconstruction is possible when one has access to only corrupted measurement examples and the underlying images share common, low-dimensional structure. We explore such ideas in this work in the following two ways: Left: Given a single measurement example (with the corresponding forward model) and a set of candidate image generation models (IGMs), we use our proposed criterion to select the best IGM from the measurements alone and then recover an image posterior under that IGM. The criterion we use for model selection is an approximation of the evidence lower bound (i.e., ELBOProxy in (3)) (see Section III-A). Right: Building off of this, we consider the setting where we must infer the IGM from noisy measurements alone (see Sec. III-B). In this case, we solve N inverse problems simultaneously. The inputs of this method are N measurement examples with their known forward models, and the outputs are a single inferred IGM and N latent embeddings that, when combined, lead to N image reconstruction posteriors.

of highest likelihood. Unfortunately computing $p(y|G)$ is intractable, as it requires marginalizing and integrating over all x encompassed by the IGM G. However, we show that this quantity can be well approximated using the ELBO.

To motivate our discussion, we first consider estimating the image posterior $p(x|y, G)$ by learning the parameters ϕ of a variational distribution $h_\phi(x)$. Observe that the definition of the KL-divergence followed by an application of Bayes' theorem gives

$$D_{\mathrm{KL}}\left(h_\phi(x) \parallel p(x|y, G)\right) := \mathbb{E}_{x \sim h_\phi(x)}\left[\log \frac{h_\phi(x)}{p(x|y, G)}\right]$$

$$= \mathbb{E}_{x \sim h_\phi(x)}\left[\log \frac{h_\phi(x)p(y|G)}{p(y|x, G)p(x|G)}\right]$$

$$= -\mathbb{E}_{x \sim h_\phi(x)}[\log p(y|x, G) + \log p(x|G) - \log h_\phi(x)]$$

$$+ \log p(y|G).$$

The ELBO of an IGM G given measurements y under variational distribution h_ϕ is defined by

$$\mathrm{ELBO}(G, h_\phi; y) := \mathbb{E}_{x \sim h_\phi(x)}[\log p(y|x, G)$$

$$+ \log p(x|G) - \log h_\phi(x)]. \quad (2)$$

Rearranging the previous equation, we see that by the non-negativity of the KL-divergence that

$$\log p(y|G) = D_{\mathrm{KL}}\left(h_\phi(x) \parallel p(x|y, G)\right)$$

$$+ \mathrm{ELBO}(G, h_\phi; y)$$

$$\geqslant \mathrm{ELBO}(G, h_\phi; y).$$

Thus, we can lower bound the model posterior as

$$\log p(G|y) \geqslant \mathrm{ELBO}(G, h_\phi; y) + \log p(G) - \log p(y).$$

Note that $\log p(y)$ is independent of the parameters of interest, ϕ. If the variational distribution $h_\phi(x)$ is a good approximation to the posterior $p(x|y, G)$, $D_{\mathrm{KL}} \approx 0$. Thus, maximizing $\log p(G|y)$ with respect to G is approximately equivalent to maximizing $\mathrm{ELBO}(G, h_\phi; y) + \log p(G)$.

Each term in the ELBO objective encourages certain properties of the IGM G. In particular, the first term in the ELBO, $\mathbb{E}_{x \sim h_\phi(x)}[\log p(y|x, G)]$, requires that G should lead to an image estimate that is consistent with our measurements y. The second term, $\mathbb{E}_{x \sim h_\phi(x)}[\log p(x|G)]$, encourages images sampled from $h_\phi(x)$ to have high likelihood under our model G. The final term is the entropy term, $\mathbb{E}_{x \sim h_\phi(x)}[-\log h_\phi(x)]$, which encourages a G that leads to "fatter" minima that are less sensitive to small changes in likely images x under G.

1) ELBOProxy: Some IGMs are explicit, which allows for direct computation of $\log p(x|G)$. For example, if our IGM models x as isotropic Gaussian with variance λ, then $-\log p(x|G) \propto \lambda^{-1}\|x\|_2^2$. In this case, we can optimize the ELBO defined in Equation (2) directly and then perform model selection. However, an important class of IGMs that we are interested in are those given by deep generative networks. Such IGMs are not probabilistic in the usual Bayesian interpretation of a prior, but instead implicitly enforce structure in the data. A key characteristic of many generative network architectures (e.g., VAEs and GANs) that we leverage is that they generate high-dimensional images from low-dimensional latent representations. Bottlenecking helps the network learn global characteristics of the underlying image distribution while also respecting the low intrinsic dimensionality of natural images. However, this means that we can only compute $\log p(x|G)$ directly if we have an injective map [29]. This architectural requirement limits the expressivity of the network.

We instead consider a proxy of the ELBO that is especially helpful for deep generative networks. That is, suppose our IGM is of the form $x = G(z)$. Introducing a variational family for our latent representations $z \sim q_\phi(z)$ and choosing a latent prior distribution $\log p_Z(z|G)$, we arrive at the following proxy of the ELBO:

$$\mathrm{ELBOProxy}(G, q_\phi; y) := \mathbb{E}_{z \sim q_\phi(z)}[\log p(y|G(z))$$

$$+ \log p_Z(z|G) - \log q_\phi(z)]. \quad (3)$$

In our experiments, we chose $p_Z(z|G)$ to be an isotropic Gaussian prior. This is a common choice in many generative modeling

frameworks and has shown to be a good choice of prior in the latent space.

To motivate this proxy, it is instructive to consider the case where our variational distribution $h_\phi := G \sharp q_\phi$ is the push-forward of a latent distribution q_ϕ through an injective or invertible function G. To be precise, recall the following definition of the push-forward measure.

Definition 1: Let $G : \mathbb{R}^k \to \mathbb{R}^n$ be a measurable function and suppose p is a distribution (or, more generally, a measure) on $\mathbb{R}^k$. Then the push-forward measure $\mu := G \sharp p$ is the measure on $\mathbb{R}^n$ that satisfies the following: for all Borel sets A of $\mathbb{R}^n$, $\mu(A) = p(G^{-1}(A))$ where $G^{-1}(A)$ denotes the preimage of A with respect to G.

The push-forward measure essentially characterizes how a distribution p changes when passed through a function G. It follows from the definition of the push-forward that $x \sim G \sharp q_\phi$ if and only if $x = G(z)$ where $z \sim q_\phi$. In the case $h_\phi = G \sharp q_\phi$ for an injective function G, the ELBO and ELBOProxy are equivalent, as shown in the following proposition:

Proposition 1. Suppose $G : \mathbb{R}^k \to \mathbb{R}^n$ is continuously differentiable and injective. For two probability distributions p_Z and q_ϕ on $\mathbb{R}^k$, define the measures $p(\cdot|G) = G \sharp p_Z$ and $h_\phi = G \sharp q_\phi$. Then

$$\text{ELBO}(G, h_\phi; y) = \text{ELBOProxy}(G, q_\phi; y) \; \forall y \in \mathbb{R}^m.$$

Proof: It suffices to show

$$\mathbb{E}_{x \sim h_\phi(x)}[\log p(x|G) - \log h_\phi(x)]$$
$$= \mathbb{E}_{z \sim q_\phi(z)}[\log p_Z(z|G) - \log q_\phi(z)]$$

Let $J_G(z) \in \mathbb{R}^{n \times k}$ denote the Jacobian of G at an input $z \in \mathbb{R}^k$. Since G is injective and continuously differentiable with $p(\cdot|G) = G \sharp p_Z$, we can compute the likelihood of any point $x \in \text{range}(G)$ [29] via

$$\log p(x|G) = \log p_Z(G^\dagger(x)|G)$$
$$- \frac{1}{2} \log |\det[J_G(G^\dagger(x))^T J_G(G^\dagger(x))]|$$

where $G^\dagger$ is the inverse of G on its range. This is essentially the classical change-of-variables formula specialized to the case when G is injective and we wish to access likelihoods on the range of G. Note that this equation is only valid for points in the range of the injective function G. Likewise, since $h_\phi = G \sharp q_\phi$, we can compute the entropy of h_ϕ for any point $x \in \text{range}(G)$ as

$$\log h_\phi(x) = \log q_\phi(G^\dagger(x))$$
$$- \frac{1}{2} \log |\det[J_G(G^\dagger(x))^T J_G(G^\dagger(x))]|.$$

Now observe that for $x \sim h_\phi$, $x \in \text{range}(G)$ as h_ϕ is the push-forward of q_ϕ through G. Thus, for $x \sim h_\phi$, we have that

$$\log p(x|G) - \log h_\phi(x)$$
$$= \log p_Z(G^\dagger(x)|G) - \log q_\phi(G^\dagger(x)).$$

By the definition of the push-forward measure, we have that $x \sim h_\phi$ implies $x = G(z)$ for some $z \sim q_\phi$. Using our previous formulas, we can compute the expectation over the difference $\log p(x|G) - \log h_\phi(x)$ with respect to h_ϕ as

$$\mathbb{E}_{x \sim h_\phi(x)}[\log p(x|G) - \log h_\phi(x)]$$
$$= \mathbb{E}_{x \sim h_\phi(x)}[\log p_Z(G^\dagger(x)|G) - \log q_\phi(G^\dagger(x))]$$
$$= \mathbb{E}_{z \sim q_\phi(z)}[\log p_Z(G^\dagger(G(z))|G) - \log q_\phi(G^\dagger(G(z)))]$$
$$= \mathbb{E}_{z \sim q_\phi(z)}[\log p_Z(z|G) - \log q_\phi(z)].$$

□

An important consequence of this result is that for injective generators G, the inverse of G (on its range) is not required for computing the ELBO. In this case, the ELBOProxy is in fact equivalent to the ELBO. While not all generators G will be injective, quality generators are largely injective over high likelihood image samples. In Section III-A2 and Fig. 3, we experimentally show that this proxy can aid in selecting potentially non-injective generative networks from corrupted measurements.

2) Toy Example: To illustrate the use of the ELBOProxy as a model selection criterion, we conduct the following experiment that asks whether the ELBOProxy can identify the best model from a given set of image generation models. For this experiment, we use the MNIST dataset [32] and consider two inverse problems: denoising and phase retrieval. We train a generative model G_c on each class $c \in \{0, 1, 2, \ldots, 9\}$ using the clean MNIST images directly. Hence, G_c generates images from class c via $G_c(z)$ where $z \sim \mathcal{N}(0, I)$. Then, given noisy measurements y_c from a single image from class c, we ask whether the generative model G_c from the appropriate class would achieve the best ELBOProxy. Each G_c is the decoder of a VAE with a low-dimensional latent space, with no architectural constraints to ensure injectivity. For denoising, our measurements are $y_c = x_c + \eta_c$ where $\eta_c \sim \mathcal{N}(0, \sigma^2 I)$ and $\sigma = \sqrt{0.5}$. For phase retrieval, $y_c = |\mathcal{F}(x_c)| + \eta_c$ where $\mathcal{F}$ is the Fourier transform and $\eta_c \sim \mathcal{N}(0, \sigma^2 I)$ with $\sigma = \sqrt{0.05}$.

We construct 10×10 arrays for each problem, where in the i-th row and j-th column, we compute the negative ELBOProxy obtained by using model G_{i-1} to reconstruct images from class $j - 1$. We calculate $\text{ELBOProxy}(G_c, q_{\phi_c}; y_c)$ by parameterizing q_{ϕ_c} with a Normalizing Flow [18] and optimizing network weights ϕ_c to maximize (3). The expectation in the ELBOProxy is approximated via Monte Carlo sampling. Results from the first 5 classes are shown in Fig. 3 and the full arrays are shown in the supplemental materials. We note that all of the correct models are chosen in both denoising and phase retrieval. We also note some interesting cases where the ELBOProxy values are similar for certain cases, such as when recovering the 3 or 4 image. For example, when denoising the 4 image, both G_4 and G_9 achieve comparable ELBOProxy values. By carefully inspecting the noisy image of 4, one can see that both models are reasonable given the structure of the noise.

B. Simultaneously Solving Many Inverse Problems

As the previous section illustrates, the ELBOProxy provides a good criterion for choosing an appropriate IGM from noisy measurements. Here, we consider the task of directly inferring
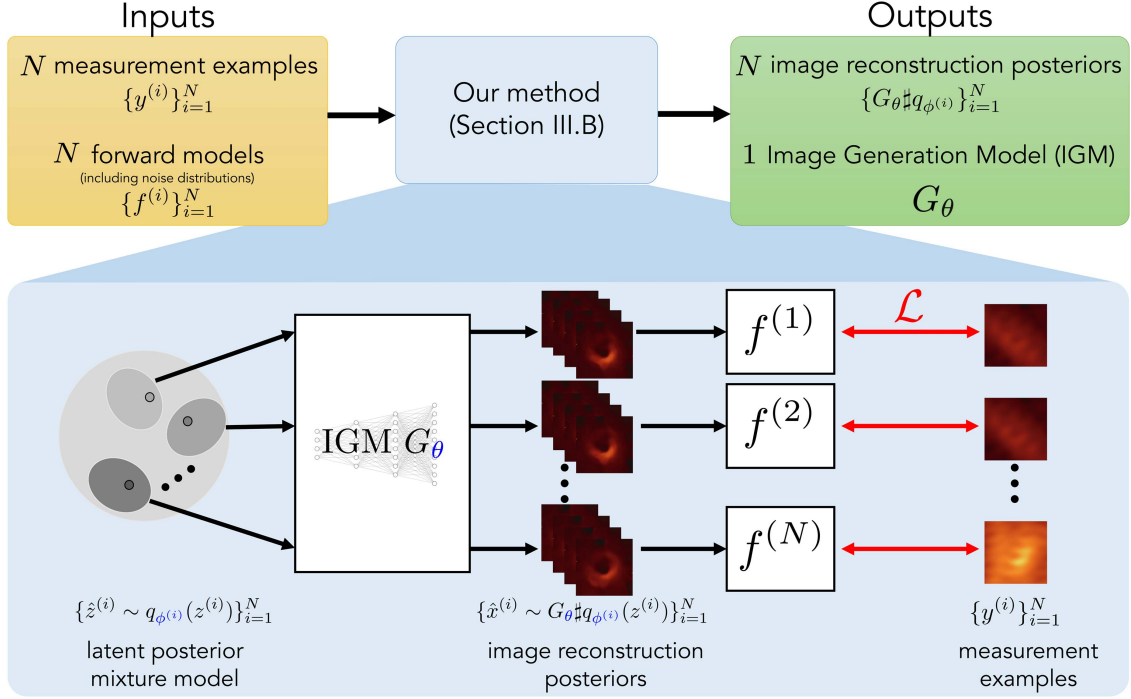
Fig. 2. We propose to solve a collection of N ill-posed inverse problems by exploiting the common, low-dimensional structure of the underlying images. Given a set of N measurement examples $\{y^{(i)}\}_{i=1}^N$ from N different underlying images, we propose to model each image posterior as the output of a shared IGM with a low-dimensional latent space. In particular, each posterior is approximated by $G_\theta \sharp q_{\phi^{(i)}}$, the push-forward of $q_{\phi^{(i)}}$ through G_θ, where G_θ is the shared, common generator to all N examples and $q_{\phi^{(i)}}$ is a low-dimensional, variational distribution particular to the i-th example. The parameters, θ and $\{\phi^{(i)}\}_{i=1}^N$, are colored in blue and are jointly inferred. That is, such parameters are inferred specifically for the measurements $\{y^{(i)}\}_{i=1}^N$. The loss we use is the negative ELBOProxy, which is denoted by $\mathcal{L}$ and given by (5). Note that there is no notion of a training set and test set as we aim to solve the inverse problems jointly from all available measurements.

the IGM from a collection of measurement examples $y^{(i)} = f^{(i)}(x^{(i)}) + \eta^{(i)}$ for $i \in [N]$, where the parameters are found by optimizing the ELBOProxy. The key assumption we make is that common, low-dimensional structure is shared across the underlying images $\{x^{(i)}\}_{i=1}^N$. We propose to find a *shared* generator G_θ with weights θ along with latent distributions $q_{\phi^{(i)}}$ that can be used to reconstruct the full posterior of each image $x^{(i)}$ from its corresponding measurement example $y^{(i)}$. This approach is illustrated in Fig. 2. Having the generator be shared across all images helps capture their common collective structure. Each forward model corruption, however, likely induces its own complicated image posteriors. Hence, we assign each measurement example $y^{(i)}$ its own latent distribution to capture the differences in their posteriors. Note that because we optimize a proxy of the ELBO, the inferred distribution may not necessarily be the true image posterior, but it still captures a distribution of images that fit to the observed measurements.

a) Inference approach: More explicitly, given a collection of measurement examples $\{y^{(i)}\}_{i=1}^N$, we jointly infer a generator G_θ and a set of variational distributions $\{q_{\phi^{(i)}}\}_{i=1}^N$ by optimizing a Monte Carlo estimate of the ELBOProxy from (3), described by:

$$\{\hat{\theta}, \hat{\phi}^{(1)}, \dots, \hat{\phi}^{(N)}\} \in \arg\max_{\theta, \{\phi^{(i)}\}_{i=1}^N} \mathcal{L} \qquad (4)$$

where

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \text{ELBOProxy}(G_\theta, q_{\phi^{(i)}}; y^{(i)}) + \log p(G_\theta). \quad (5)$$

In terms of choices for $\log p(G_\theta)$, we can add additional regularization to promote particular properties of the IGM G_θ, such as having a small Lipschitz constant. Here, we consider having sparse neural network weights as a form of regularization and use dropout [49] during training to represent $\log p(G_\theta)$.

Once a generator $G_{\hat{\theta}}$ and variational parameters $\hat{\phi}^{(i)}$ have been inferred, we solve the i-th inverse problem by simply sampling $\hat{x}^{(i)} = G_{\hat{\theta}}(\hat{z}^{(i)})$ where $\hat{z}^{(i)} \sim q_{\hat{\phi}^{(i)}}(z^{(i)})$ or computing an average $\overline{x}^{(i)} = \frac{1}{T} \sum_{t=1}^T G_{\hat{\theta}}(\hat{z}_t^{(i)})$. Producing samples for each inverse problem can help visualize the range of uncertainty under the learned IGM $G_{\hat{\theta}}$, while the expected value of the distribution empirically provides clearer estimates with better metrics in terms of PSNR or MSE. We report PSNR outputs in our subsequent experiments and also visualize the standard deviation of our reconstructions.

IV. EXPERIMENTAL RESULTS

We now consider solving a set of inverse problems via the framework described in III-B. For each of these experiments, we use a multivariate Gaussian distribution to parameterize each
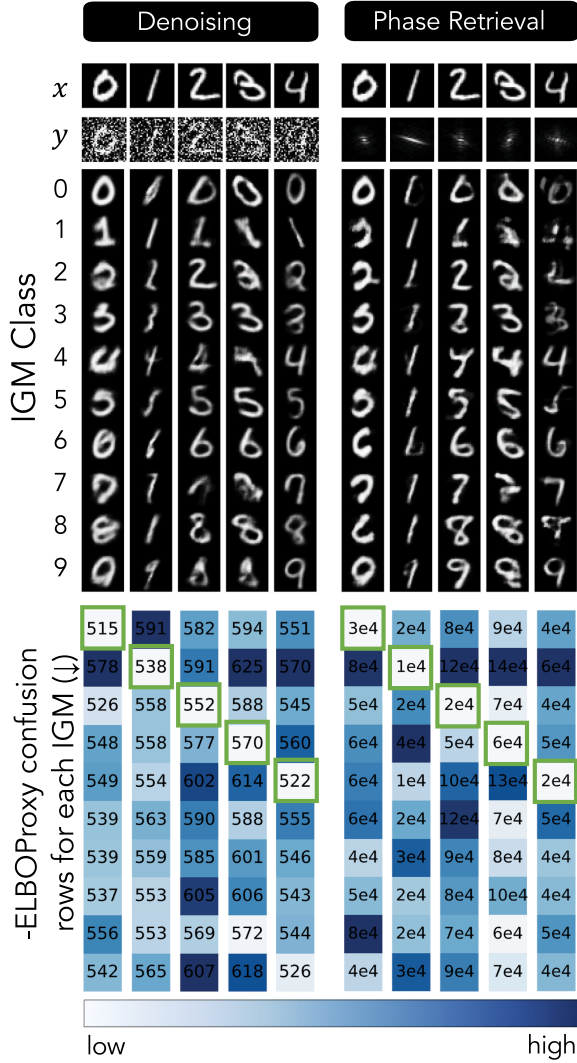
Fig. 3. We consider two inverse problems: denoising and phase retrieval. Top: the two topmost rows correspond to the ground-truth image x_c and the noisy measurements y_c. Center: in each column, we show the means of the distribution induced by the push-forward of G_j and each latent distribution $z \sim q_{\phi_j}$ for $j \in \{0, \ldots, 9\}$. Bottom: each column of the array corresponds to the negative ELBOProxy achieved by each model in reconstructing the images. Here, lower is better. Boxes highlighted in green correspond to the best negative ELBOProxy values in each column. In all these examples, the correct model was chosen.

of the posterior distributions $q_{\phi^{(i)}}$ and a Deep Decoder [24] with 6 layers, 150 channels in each layer, a latent size of 40, and a dropout of 10^{-4} as the IGM. The multivariate Gaussian distributions are parameterized by means and covariance matrices $\{\mu^{(i)}, \Lambda^{(i)} = U_i U_i^T + \varepsilon I\}_{i=1}^N$, where εI with $\varepsilon = 10^{-3}$ is added to the covariance matrix to help with stability of the optimization. We choose to parameterize the latent distributions using Gaussians for memory considerations. Note that the same hyperparameters are used for all experiments demonstrating our proposed method.

In our experiments, we also compare to the following baseline methods: AmbientGAN [5], Deep Image Prior (DIP) [54], and regularized maximum likelihood using total variation (TV-RML). AmbientGAN is most similar to our setup, as it

constructs an IGM directly from measurement examples; however, it doesn't aim to estimate image reconstruction posteriors, but instead aims to learn an IGM that samples from the full underlying prior. TV-RML uses an explicit total variation regularization, while DIP uses implicit convolutional neural network "prior". As we will show, all these baseline methods require fine-tuning hyperparameters to each set of measurements in order to produce their best results. All methods can also be applied to a variety of inverse problems, making them appropriate choices as baselines.

A. Denoising

We show results on denoising a collection of noisy images of 8's from the MNIST dataset in Fig. 4 and denoising a collection of noisy images of a single face from the PubFig [31] dataset in Fig. 5. The measurements for both datasets are defined by $y = x + \eta$ where $\eta \sim \mathcal{N}(0, \sigma^2 I)$ with an SNR of $\sim$-3 dB for the MNIST digits and an SNR of $\sim$15 dB for the faces. Our method is able to remove much of the added noise and recovers small scale features, even with only 10's of observations. As shown in Fig. 4, the reconstructions achieved under the learned IGM improves as the number of independent observations increases. Our reconstructions also substantially outperform the baseline methods, as shown in Fig. 6. Unlike DIP, our method does not overfit and does not require early stopping. Our method does not exhibit noisy artifacts like those seen in all baselines methods, despite such methods being fine-tuned. We show quantitative comparisons in Table I.

In Fig. 7 we show additional *multi-noise* denoising experiments where we have 75 noisy images, which have 3 different noise levels. More formally, $y^{(i)} = x^{(i)} + \eta$ where $\eta \in \{\eta_1, \eta_2, \eta_3\}$ and $\eta_i \sim \mathcal{N}(0, \sigma_i^2 I)$. In Experiment 1, the noise levels have a wide range, and we use standard deviations of $\{\sigma_1, \sigma_2, \sigma_3\} = \{0.01, 0.1, 0.5\}$. In Experiment 2, the noise levels are much closer together, and we use standard deviations of $\{\sigma_1, \sigma_2, \sigma_3\} = \{0.2, 0.3, 0.5\}$. When the SNRs are similar (as in Experiment 2), the reconstructions match the true underlying images well. However, when the measurements have a wide range of SNRs (i.e., Experiment 1), the reconstructions from low SNR measurements exhibit bias and poorly reconstruct the true underlying image, as shown in Fig. 7. This is likely because the high SNR measurements influence the inferred IGM more strongly than the low SNR measurements. The full set of results are available in the supplemental materials.

B. Phase Retrieval

Here we consider solving non-convex inverse problems, and demonstrate our approach on phase retrieval. Our measurements are described by $y = |\mathcal{F}(x)| + \eta$ where $\mathcal{F}(\cdot)$ is a linear operator and $\eta \sim \mathcal{N}(0, \sigma^2 I)$. We consider two types of measurements, one where $\mathcal{F}(\cdot)$ is the Fourier transform and the other when $\mathcal{F}(\cdot)$ is an $m \times n$ complex Gaussian matrix with $m = \lceil 0.1n \rceil$. Since each measurement is the magnitude of complex linear measurements, there is an inherent phase ambiguity in the problem. Additionally, flipping and spatial-shifts are possible reconstructions when performing Fourier phase retrieval. Due to
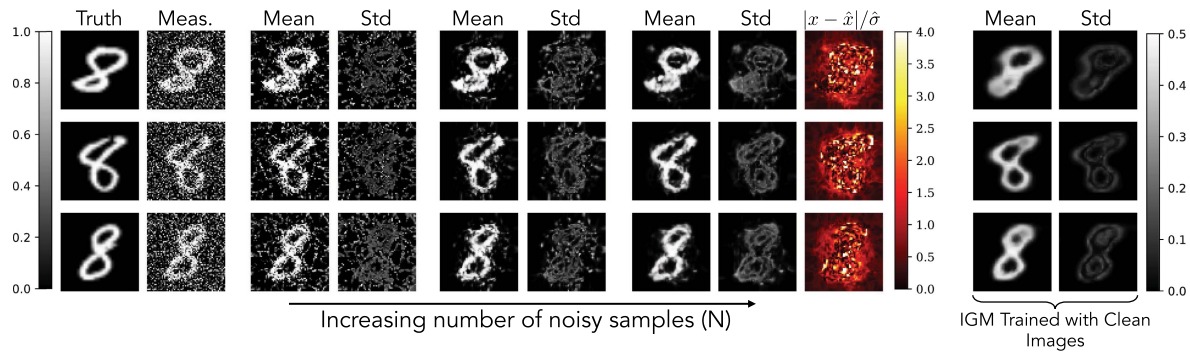
Fig. 4. Improvement with an increasing number of noisy observations. We demonstrate our method of inferring an IGM to perform denoising for an increasing number of noisy MNIST images (5, 75, and 150 images from left to right). We showcase results on three randomly selected examples that appear in each collection of inverse problems. In each panel, we include the ground-truth, noisy measurements, mean of the posterior, and standard deviation of the posterior. We also include the residual error divided by the empirical standard deviation for $N = 150$. On the far right, we visualize reconstructions using an IGM trained on the full clean MNIST 8's class (6000 images). We observe that the mean reconstructions and standard deviations from our low-data IGMs become more similar to the full-data IGM with increasing data. Our residual errors are largely within 3 standard deviations.
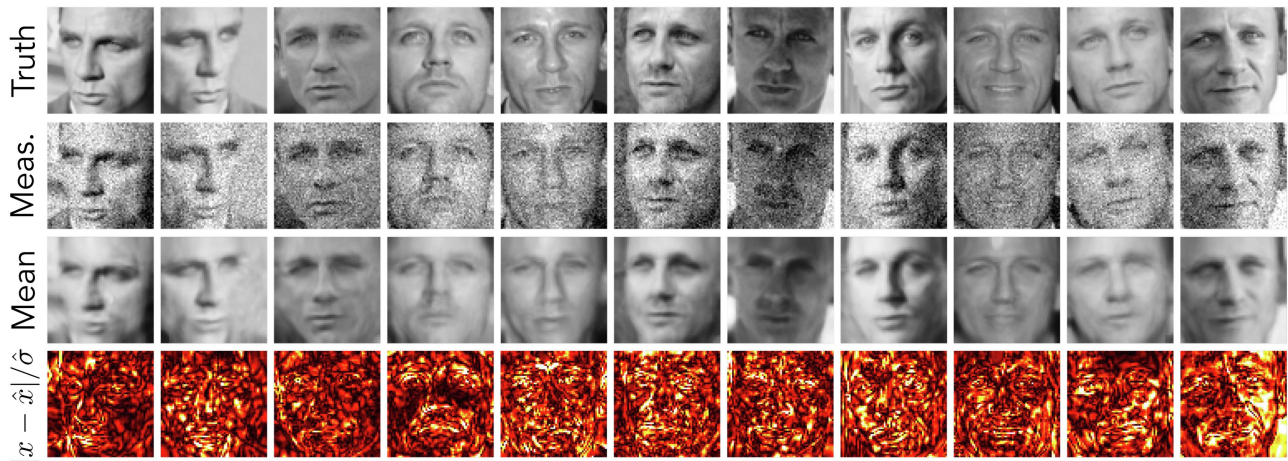

Fig. 5. Denoising 95 images of celebrity A. We demonstrate our method described in III-B using 95 noisy images of a celebrity. Here we show the underlying image (row 1), noisy measurements (row 2), mean reconstruction (row 3), and residual error (row 4) for a subset of the 95 different noisy images. Our reconstructions are much less noisy than the measurements and recover sharp features that are hard to discern in the noisy images. We visualize the residual error normalized by the empirical standard deviation, which indicates errors are largely within 3 standard deviations (refer to the colorbar in Fig. 4). Note that no explicit spatial-domain prior/regularizer was used in denoising.
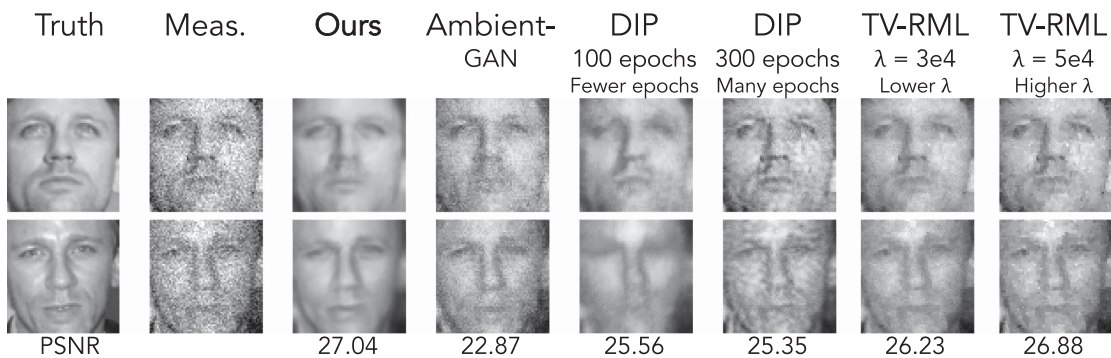
Fig. 6. Denoising baseline comparisons. We compare to various baselines (AmbientGAN, Deep Image Prior (DIP), and regularized maximum likelihood using TV (TV-RML) with weight λ), and we report the average PSNR across all 95 reconstructions. We show both early stopping and full training results using DIP. Our method exhibits higher PSNR than all other baselines. We also include results for baselines that require fine-tuning to demonstrate sensitivity to subjective stopping conditions.

TABLE I
QUANTITATIVE COMPARISONS TO BASELINES

Dataset	N	Forward model $y = f(x) + \eta$	Ours	AmbGAN fine-tuned	DIP fewer epochs	DIP many epochs	TV lower λ	TV higher λ
						Baselines		
celeb. A	95	$y = x + \eta$	**27.0**	22.9	25.6	25.4	26.2	26.9
celeb. B	95	$y = x + \eta$	26.2	19.7	24.4	25.2	25.9	**26.4**
MNIST 8's	150	$y = x + \eta$	**21.1**	18.0	18.8	13.3	16.4	18.2
M87* (target)	60	$y = Ax + \eta$	**29.3**	25.7	29.0	28.6	24.3	25.9

We show the mean PSNR between the reconstructions of different methods (ours, AmbientGAN [5], DIP [54], and TV-RML) with either the true underlying image or the target in the case of black hole compressed sensing. The highest PSNR for each set of measurements is in bold. We fix the hyperparameters of the baselines across each forward model, empirically selecting them for good performance on each type of problem. Note that the range in performance depends on the choice of hyperparameters. For denoising, the DIP results corresponds to 100 and 300 epochs for the fewer and many epochs baselines, respectively, and the lower and higher λ's are 1e3 and 1e4, respectively. For black hole compressed sensing, the DIP results corresponds to 1000 and 3000 epochs for the fewer and many epochs baselines, respectively, and the lower and higher λ's are 1e2 and 1e3, respectively. Visual examples of celeb B. are shown in the supplemental materials.
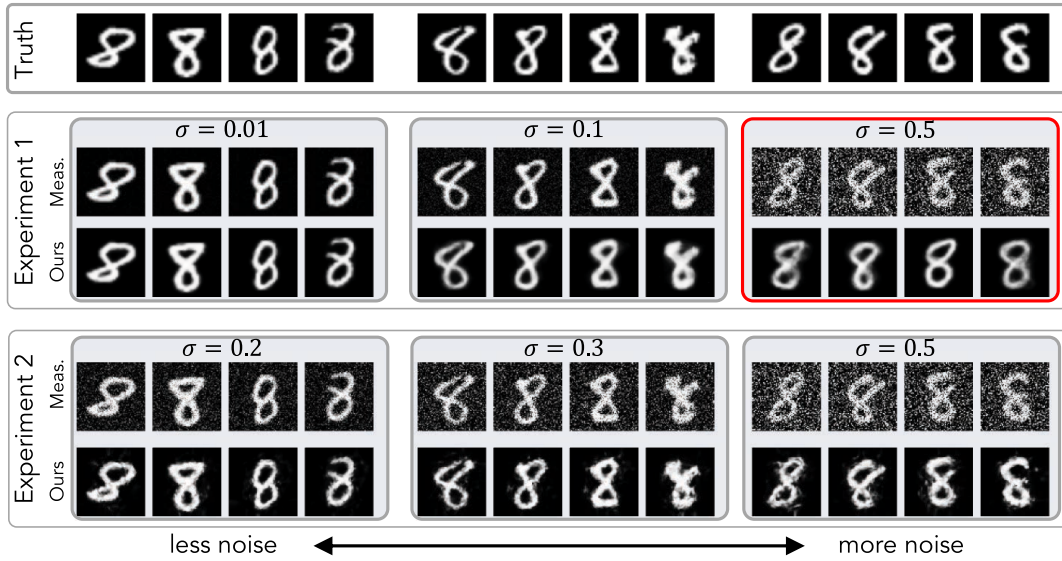

Fig. 7. Multi-noise denoising. We demonstrate our method described in Section III-B to perform denoising on measurement collections experiencing different noise levels. For each experiment, we use 75 measurement examples, which are defined by $y^{(i)} = x^{(i)} + \eta$ where $\eta \in \{\eta_1, \eta_2, \eta_3\}$ and $\eta_i \sim \mathcal{N}(0, \sigma_i^2 I)$. In Experiment 1, we use noisy measurement examples that have additive noise with standard deviations of 0.01, 0.1, and 0.5. In Experiment 2, we use noisy measurement examples that have additive noise with standard deviations of 0.2, 0.3, and 0.5. We visualize the true underlying images, the measurement used for each experiment, and the mean of the image reconstruction posterior. Most of the reconstructions recover the primary features of the true image. However, in Experiment 1, the reconstructions of the low SNR measurements exhibit bias and do not match the true images well. This is likely because in Experiment 1, the high SNR measurements influence the inferred IGM more strongly than low SNR measurements, leading to biased reconstructions for the reconstructions highlighted in the red box.

the severe ill-posedness of the problem, representing this complicated posterior that includes all spatial shifts is challenging. Thus, we incorporate an envelope (i.e., a centered rectangular mask) as the final layer of G_θ to encourage the reconstruction to be centered. Nonetheless, flipping and shifts are still possible within this enveloped region.

We show results from a set of $N = 150$ noisy phase retrieval measurements from the MNIST 8's class with a SNR of $\sim$52 dB. We consider three settings: 1) all measurements arise from a Gaussian measurement matrix, 2) all measurements arise from Fourier measurements, and 3) half of the measurements are Gaussian and the other half are Fourier. We show qualitative results for cases 1 and 2 in Fig. 8. In the Gaussian case, we note that our mean reconstructions are nearly identical to the true digits and the standard deviations exhibit uncertainty in

regions we would expect (e.g., around edges). In the Fourier case, our reconstructions have features similar to the digit 8, but contain artifacts. These artifacts are only present in the Fourier case due to additional ambiguities, which lead to a more complex posterior [27]. We also show the average PSNR of our reconstructions for each measurement model in Table II. For more details on this experiment, please see Section XII-C of the supplemental materials.

C. Black Hole Imaging

We consider a real-world inverse problem for which ground-truth data would be impossible to obtain. In particular, we consider a compressed sensing problem inspired by astronomical imaging of black holes with the Event Horizon Telescope
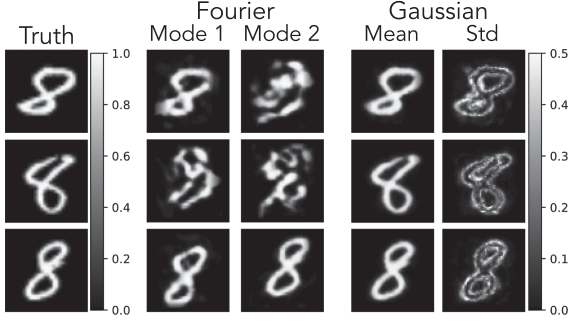
Fig. 8. Phase retrieval from MNIST 8's. We demonstrate our method described in III-B to perform phase retrieval on 150 images. For the Fourier phase retrieval setting, we show examples from the two observed modes of the posterior. For the Gaussian phase retrieval setting, we show the mean and standard deviation of our reconstructions.

TABLE II
PHASE RETRIEVAL PSNRs FOR DIFFERENT MEASUREMENT OPERATORS

	Measurement Operator(s)		
	Gaussian	Fourier	Both
Gaussian	30.8	—	30.2
Fourier	—	13.6	19.4

Each column corresponds to the type of measurements that the method was given (a total of $N = 150$). In the case of "Both", 75 Gaussian measurements and 75 Fourier measurements were given. We then show the average PSNR for our reconstructions given the specific measurement operator (either Fourier or Gaussian). Note that when given both Gaussian and Fourier measurement examples our method exhibits a higher PSNR on the recovered images from Fourier examples than when given only Fourier measurements. Additionally, there is only a slight decrease in performance on the Gaussian measurement reconstructions as compared to given entirely Gaussian examples.

(EHT): suppose we are given access to N measurement examples of the form $y^{(i)} = A^{(i)}x^{(i)} + \eta^{(i)}$ where $A^{(i)} \in \mathbb{C}^{m \times n}$ is a low-rank compressed sensing matrix arising from interferometric telescope measurements and $\eta^{(i)}$ denotes noise with known properties (e.g., distributed as a zero-mean Gaussian with known covariance). The collection of images $\{x^{(i)}\}_{i=1}^{N}$ are snapshots of an evolving black hole target. This problem is ill-posed and requires the use of priors or regularizers to recover a reasonable image [13]. Moreover, it is impossible to directly acquire example images of black holes, so any pixel-level prior defined *a priori* will exhibit human bias. Recovering an image, or movie, of a black hole with as little human bias as possible is essential for both studying the astrophysics of black holes as well as testing fundamental physics [12], [15]. We show how our proposed method can be used to tackle this important problem. In particular, we leverage knowledge that, although the black hole evolves, it will not change drastically from minute-to-minute or day-to-day. We study two black hole targets: the black hole at the center of the Messier 87 galaxy (M87*) and the black hole at the center of the Milky Way galaxy – Sagattarius A* (Sgr A*).
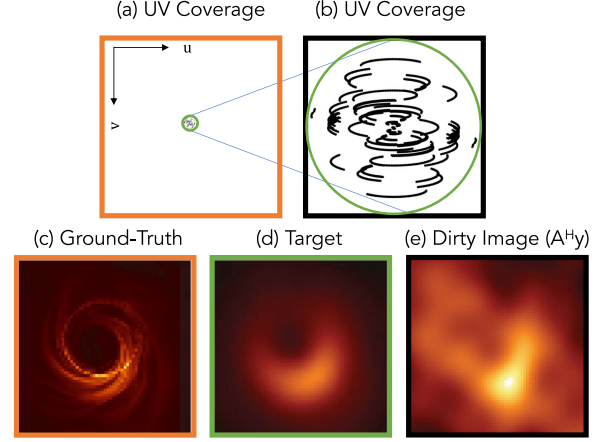

Fig. 9. Visualization of the intrinsic resolution of the EHT compressed sensing measurements. The EHT measures sparse spatial frequencies of the image (i.e., components of the image's Fourier transform). In order to generate the underlying image (c), all frequencies in the entire domain of (a) must be used. Restricting spatial frequencies to the ones in (a) and (b)'s green circle generates the target (d), where (b) is a zoom in of (a). The EHT samples a subset of the interior of the green circle, indicated by the sparse black samples in (b). Naively recovering an image using only these frequencies results in the *dirty image* (e), which is computed by $A^H y$. The 2D spatial Fourier frequency coverage represented with (u, v) positions is referred to as the UV coverage.

a) Imaging M87* using the current EHT array: We first consider reconstructing the black hole at the center of the Messier 87 galaxy, which does not evolve noticeably within the timescale of a single day. The underlying images are from a simulated 60 frame video with a single frame for each day. We show results on $N = 60$ frames from an evolving black hole target with a diameter of $\sim$40 microarcseconds, as was identified as the diameter of M87 according to [14], [42] in Fig. 10. In particular, the measurements are given by $\{y^{(i)} = Ax^{(i)} + \eta^{(i)}\}_{i=1}^{N}$, where $x^{(i)}$ is the underlying image on day i, A is the forward model that represents the telescope array, which is static across different days, and the noise $\eta^{(i)} \sim \mathcal{N}(0, \Sigma)$ has a covariance of Σ, which is a diagonal matrix with realistic variances based on the telescope properties.[1] Measurements are simulated from black hole images with a realistic flux of 1 Jansky [57]. We also visualize a reference "target" image, which is the underlying image filtered with a low-pass filter that represents the maximum resolution achievable with the telescope array used to collect measurements – in this case the EHT array consisting of 11 telescopes (see Fig. 9 and Section IX in the supplemental materials).

As seen in Fig. 10, our method is not only able to reconstruct the large scale features of the underlying image without any aliasing artifacts, but also achieves a level of super-resolution (see Table III in the supplemental materials). Our reconstructions also achieve higher super-resolution as compared to our baselines (i.e., AmbientGAN, TV-RML, and DPI) in Fig. 10 and do not exhibit artifacts evident in the reconstructions from these baselines. The two AmbientGAN settings were qualitatively chosen to show that the final result is sensitive to the

[1]We leave the more challenging atmospheric noise that appears in measurements for future work.
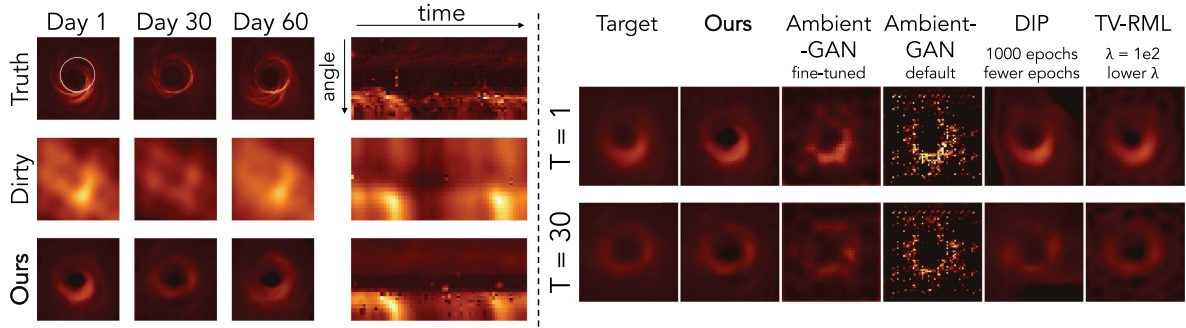
Fig. 10. Recovering a 60-day video of the M87 black hole with the EHT telescope array. Left: We demonstrate our method described in III-B using simulated measurements from 60 frames of an evolving black hole target with the forward model described in Fig. 9. Here we show the underlying images, dirty images ($A^H y$, see Fig. 9), and mean reconstruction, respectively. Additionally, we show the unwrapped space × time image, which is taken along the overlaid white ring illustrated in the day 1 underlying image. The bright-spot's temporal trajectory of our reconstruction matches that of the truth. Right: We compare our method to various baselines methods. Our results are much sharper and exhibit less artifacts than AmbientGAN and TV-RML with weight λ. Note that we include results using AmbientGAN with both default parameters and fine-tuned parameters.

choice of hyperparameters. The default AmbientGAN parameters produce poor results, and even with fine-tuning to best fit the underlying images (i.e., cheating with knowledge of the ground-truth), the results still exhibit substantial artifacts. We outperform the baselines in terms of PSNR when compared to the target image (see Table I). Our results demonstrate that we are able to capture the varying temporal structure of the black hole, rather than just recovering a static image. It is important to note that there is no explicit temporal regularization introduced; the temporal smoothness is implicitly inferred by the constructed IGM.

b) Imaging Sgr A* from multiple forward models: The framework we introduce can also be applied to situations in which the measurements themselves are induced by different forward models. In particular, the measurements $\{y^{(i)} = f^{(i)}(x^{(i)}) + \eta^{(i)}\}_{i=1}^{N}$ are given by an underlying image $x^{(i)}$, a forward model $f^{(i)}$ that is specific to that observation, and noise $\eta^{(i)}$ with known properties.

As an illustrative example, we consider the problem of reconstructing a video of the black hole at the center of the Milky Way – Sagittarius A* (Sgr A*). Unlike M87*, Sgr A* evolves on the order of minutes. Therefore, we can only consider that the black hole is static for only a short time when only a subset of telescopes are able to observe the black hole. This results in a different measurement forward model for each frame of the black hole "movie" [16]. In particular, the measurements are given by $\{y^{(i)} = A^{(i)}x^{(i)} + \eta^{(i)}\}_{i=1}^{N}$, where $x^{(i)}$ is the underlying image at time i, $A^{(i)}$ is the forward model that incorporates the telescope configuration at that time, and $\eta^{(i)} \sim \mathcal{N}(0, \Sigma^{(i)})$ is noise where $\Sigma^{(i)}$ is a diagonal matrix with realistic standard deviations derived from the telescopes' physical properties. The measurement noise is consistent with a black hole with a flux of 2 Janskys. The measurement operator is illustrated in Fig. 11.

We show examples of reconstructing 60 frames of Sgr A* with a diameter of ~50 microarcseconds using measurements simulated from a proposed future next-generation EHT (ngEHT) [43] array, which consists of 23 telescopes. These results are shown in Fig. 12. Our reconstructions remove much of the aliasing artifacts evident in the dirty images and reconstructs the
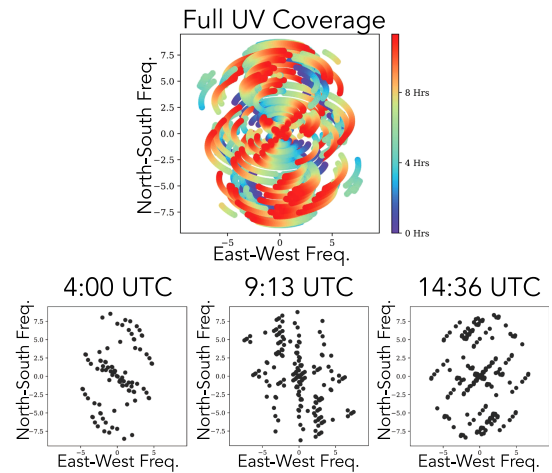

Fig. 11. Visualization of the EHT forward model of Sgr A* over time. The spatial frequency space coordinates of Sgr A* from an ngEHT [43] array measured over a course of one night is shown on the top. Snapshots of the coordinates that are measured at different times are shown on the bottom. Note that the location and sparsity of the measurement coordinates change over time.

primary features of the underlying image without any form of temporal regularization. These results have high fidelity especially considering that the measurements are very sparsely sampled. Although these results come from simulated measurements that do not account for all forms of noise we expect to encounter, the high-quality movie reconstructions obtained without the use of a spatio-temporal prior show great promise towards scientific discovery that could come from a future telescope array paired with our proposed reconstruction approach.

V. THEORY FOR LINEAR IGMS

We now introduce theoretical results on the inferred IGM for linear inverse problems. Specifically, we consider the case when the IGM G_θ is linear and the latent variational distributions are Gaussian. The goal of this section is to develop intuition for the inferred IGM in a simpler setting. While our results may not generalize to non-linear generators parameterized by deep neural
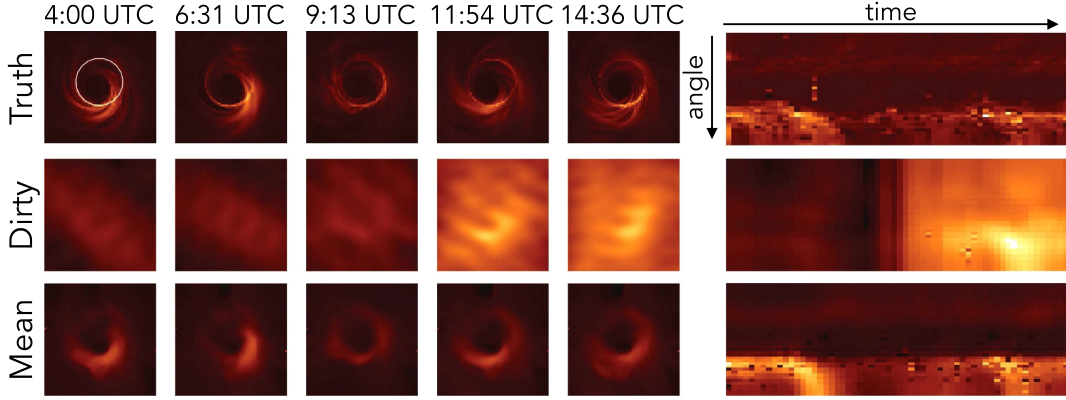
Fig. 12. Recovering a video of the Sgr A* black hole with a futuristic telescope array over the course of a night. We demonstrate our method described in III-B to perform video reconstruction on 60 frames of a video, where the measurements for each frame are generated by different forward models given by imaging Sgr A*. Here we show the underlying image, the dirty image ($A^H y$), and the mean of the reconstructed posterior. Additionally, we show the unwrapped space × time image, which is taken along the white ring illustrated in the 4:00 UTC underlying image. The bright-spot's temporal trajectory of our reconstruction matches that of the truth. The measurement noise is consistent with a black hole having a flux of 2 Janskys.

networks, our results aim to provide an initial understanding on the inferred generator.

More concretely, suppose we are given N measurement examples of noisy linear measurements of the form

$$y^{(i)} := Ax^{(i)} + \eta^{(i)}, \ \eta^{(i)} \sim \mathcal{N}(0, \sigma^2 I)$$

where $A \in \mathbb{R}^{m \times n}$ with $m \leqslant n$ and $x^{(i)} \in \mathbb{R}^n$. We aim to infer $G_\theta \in \mathbb{R}^{n \times k}$ with $k < n$ and latent distributions $q_{\phi^{(i)}} = \mathcal{N}(\mu_i, U_i U_i^T)$ where $\phi^{(i)} = \{\mu_i, U_i\}$, $\mu_i \in \mathbb{R}^k$, and $U_i \in \mathbb{R}^{k \times k}$ by minimizing the negative ELBOProxy:

$$\mathcal{L}(G_\theta, \{\phi^{(i)}\}) := \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{z \sim q_{\phi^{(i)}}(z)} [-\log p(y^{(i)} | G_\theta(z))$$
$$- \log p_Z(z | G_\theta) + \log q_{\phi^{(i)}}(z)]. \quad (6)$$

We characterize the generator G_* and latent parameters $\phi_*^{(i)} = \{\mu_i^*, U_i^*\}$ that are stationary points of (6). The result is proven in Section X:

Theorem 2: Fix $\sigma > 0$ and let $y^{(i)} = Ax^{(i)} + \eta^{(i)} \in \mathbb{R}^n$ for $i \in [N]$ where $\eta^{(i)} \sim \mathcal{N}(0, \sigma^2 I)$ and $A \in \mathbb{R}^{m \times n}$ has $k \leqslant \mathrm{rank}(A) = m \leqslant n$. Define the sample covariance of the measurements $Y_N := \frac{1}{N} \sum_{i=1}^N y^{(i)} (y^{(i)})^T$. Then, with probability 1, G_*, μ_i^*, and U_i^* that satisfy

$$AG_* := E_k (L_k - \sigma^2 I)^{1/2} R,$$

$$\mu_i^* := (G_*^T A^T A G_* + \sigma^2 I)^{-1} G_*^T A^T y^{(i)}, \text{ and}$$

$$U_i^* := \sigma (G_*^T A^T A G_* + \sigma^2 I)^{-1/2} \tilde{R},$$

for $i \in [N]$ are stationary points of the objective (6) over $\mathbb{R}^{n \times k} \times (\mathbb{R}^k \times \mathrm{GL}(k, \mathbb{R}))^N$ where $\mathrm{GL}(k, \mathbb{R})$ denotes the set of real invertible $k \times k$ matrices. Here, $E_k \in \mathbb{R}^{m \times k}$ denotes the matrix whose columns correspond to the top-$\min\{\mathrm{rank}(Y_N), k\}$ eigenvectors of Y_N, $L_k \in \mathbb{R}^{k \times k}$ contains the corresponding eigenvalues of Y_N, and $R, \tilde{R} \in \mathbb{R}^{k \times k}$ are arbitrary orthogonal matrices. If $\mathrm{rank}(Y_N) < k$, then for $\mathrm{rank}(Y_N) < i \leqslant k$, the i-th column of E_k can be arbitrary and $L_{k,ii} = \sigma^2$.

The Theorem establishes the precise form of stationary points of the objective (6). In particular, it shows that this inferred IGM performs dimensionality reduction akin to PCA [53] on the collection of measurement examples. To gain further intuition about the Theorem, in Section X-B of the supplemental materials, we analyze this result in a context where the underlying images we wish to reconstruct explicitly lie in a low-dimensional subspace. We show that, in that setting, our estimator returns an approximation of the solution found via MAP estimation, which can only be computed with complete prior knowledge of the underlying image structure. While the Theorem focused on linear IGMs, it would be interesting to theoretically analyze non-linear IGMs parameterized by deep networks. We leave this for future work.

VI. CONCLUSION

In this work we showcased how one can solve a set of inverse problems without a pre-defined IGM (e.g., a traditional spatial image prior) by leveraging common structure present across a collection of diverse underlying images. We demonstrated that even with a small set of corrupted measurements, one can jointly solve these inverse problems by directly inferring an IGM that maximizes a proxy of the ELBO. We demonstrate our method on a number of convex and non-convex imaging problems, including the challenging problem of black hole video reconstruction from interferometric measurements. Overall, our work showcases the possibility of solving inverse problems in a completely unsupervised fashion, free from significant human bias typical of ill-posed image reconstruction. We believe our approach can aid in automatic discovery of novel structure from scientific measurements, potentially paving the way to new avenues of exploration.

REFERENCES

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[2] M. Mardani et al., "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, Jan. 2019, doi: 10.1109/TMI.2018.2858752.

[3] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 524–533.

[4] A. Bora and A. Jalal, E. Price, and A. Dimakis, "Compressed sensing using generative models," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 537–546.

[5] A. Bora, E. Price, and A. G. Dimakis, "AmbientGAN: Generative models from lossy measurements," in *Proc. Int. Conf. Learn. Representations*, 2018.

[6] C. Harold, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2392–2399.

[7] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906–956, 2013.

[8] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[9] Z. Cheng, M. Gadelha, S. Maji, and D. Sheldon, "A Bayesian perspective on the deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5438–5446.

[10] B.-E. Chérief-Abdellatif, "Consistency of elbo maximization for model selection," in *Proc. 1st Symp. Adv. Approx. Bayesian Inference*, 2018, vol. 96, pp. 1–21.

[11] B.-E. Chérief-Abdellatif and P. Alquier, "Consistency of variational bayes inference for estimation and model selection in mixtures," *Electron. J. Statist.*, vol. 12, no. 2, pp. 2995–3035, 2018.

[12] The Event Horizon Telescope Collaboration, "First M87 event horizon telescope results. I. The shadow of the supermassive black hole," *Astrophysical J. Lett.*, vol. 875, no. 1, pp. L1, Apr. 2019.

[13] The Event Horizon Telescope Collaboration, "First M87 event horizon telescope results. IV. Imaging the central supermassive black hole," *Astrophysical J. Lett.*, vol. 875, no. 1, pp. L4, 2019.

[14] The Event Horizon Telescope Collaboration, "First M87 event horizon telescope results. V. Physical origin of the asymmetric ring," *Astrophysical J. Lett.*, vol. 875, no. 1, pp. L5, 2019.

[15] The Event Horizon Telescope Collaboration, "First sagittarius A* event horizon telescope results. I. the shadow of the supermassive black hole in the center of the milky way," *Astrophysical J. Lett.*, vol. 930, no. 2, pp. L12, May 2022.

[16] The Event Horizon Telescope Collaboration, "First sagittarius A* event horizon telescope results. III. imaging of the galactic center supermassive black hole," *Astrophysical J. Lett.*, vol. 930, no. 2, pp. L14, 2022.

[17] M. Z. Darestani and R. Heckel, "Accelerated MRI with un-trained neural networks," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 724–733, 2021.

[18] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real NVP," 2016, *arXiv:1605.08803*.

[19] D. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.

[20] L. Dyer, A. Parker, K. Paphiti, and J. Sanderson, "Lightsheet microscopy," *Curr. Protoc.*, vol. 2, no. 7, 2022, Art. no. e448.

[21] A. Fannjiang and T. Strohmer, "The numerics of phase retrieval," *Acta Numerica*, vol. 29, pp. 125–228, 2020.

[22] A. F. Gao, O. Leong, H. Sun, and K. L. Bouman, "Image reconstruction without explicit priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[23] P. Hand, O. Leong, and V. Voroninski, "Phase retrieval under a generative prior," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9154–9164.

[24] R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," in *Proc. Int. Conf. Learn. Representations*, 2019.

[25] R. Heckel, W. Huang, P. Hand, and V. Voroninski, "Rate-optimal denoising with deep neural networks," *Inf. Inference: J. IMA*, vol. 10, no. 4, pp. 1251–1285, 2020.

[26] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[27] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *Proc. IEEE Int. Symp. Inf. Theory Proc.*, 2013, pp. 1022–1026.

[28] M. Kabkab, P. Samangouei, and R. Chellappa, "Task-aware compressed sensing with generative adversarial networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, pp. 2297–2304, 2018.

[29] K. Kothari, A. E. Khorashadizadeh, M. d. Hoop, and I. Dokmanić, "Trumpets: Injective flows for inference and inverse problems," in *Proc. Uncertainty Artif. Intell.*, 2021, pp. 1269–1278.

[30] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void - learning denoising from single noisy images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2129–2137.

[31] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, 2009, pp. 365–372.

[32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[33] J. Lehtinen et al., "Noise2noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, vol. 80, pp. 2965–2974.

[34] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1964–1971.

[35] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," 2016, *arXiv:1606.04933*.

[36] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "RARE: Image reconstruction using deep priors learned without groundtruth," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1088–1099, Oct. 2020.

[37] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Elsevier, 1999.

[38] R. Mendoza et al., "A self-supervised approach to reconstruction in sparse x-ray computed tomography," in *Proc. NeurIPS Mach. Learn. Phys. Sci. Workshop*, 2022.

[39] S. Menon, A. Damian, N. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2437–2445.

[40] J. Miles, "R-Squared, adjusted R-squared," in *Encyclopedia of Statistics in Behavioral Science*. Hoboken, NJ, USA: Wiley, 2005.

[41] A. Olsen, Y. Hu, and V. Ganapati, "Data-driven computational imaging for scientific discovery," in *Proc. NeurIPS AI Sci. Workshop*, 2022.

[42] O. Porth et al., "The event horizon general relativistic magnetohydrodynamic code comparison project," *Astrophysical J. Suppl. Ser.*, vol. 243, no. 2, p. 26, 2019.

[43] A. W. Raymond et al., "Evaluation of new submillimeter VLBI sites for the event horizon telescope," *Astrophysical J. Suppl. Ser.*, vol. 253, no. 1, p. 5, 2021.

[44] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Med. Image Comput. Comput.- Assist. Interv.*, vol. 9351, pp. 234–241, 2015.

[46] G. E. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

[47] F. Shamshad and A. Ahmed, "Compressed sensing-based robust phase retrieval via deep generative priors," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2286–2298, Jan. 2021.

[48] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," *Int. Conf. Learn. Representations*, 2022.

[49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[50] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc.: Ser. B (Methodological)*, vol. 36, no. 2, pp. 111–133, 1974.

[51] H. Sun, K. L. Bouman, P. Tiede, J. J. Wang, S. Blunt, and D. Mawet, "Alpha-deep probabilistic inference (alpha-DPI): Efficient uncertainty quantification from exoplanet astrometry to black hole feature extraction," *Astrophysical J.*, vol. 932, no. 2, p. 99, 2022.

[52] C. Tao, L. Chen, R. Zhang, R. Henao, and L. Carin, "Variational inference and model selection with generalized evidence bounds," in *Proc. Int. Conf. Mach. Learn.*, 2018, vol. 80, pp. 893–902.

[53] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analys," *J. Roy. Stat. Soc. Ser. B: Stat. Methodol.*, vol. 61, no. 3, pp. 611–622, 1999.

[54] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9446–9454.
[55] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *Int. J. Comput. Vis.*, vol. 128, pp. 1867–1888, 2020.
[56] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.
[57] G. N. Wong et al., "PATOKA: Simulating electromagnetic observables of black hole accretion," *Astrophysical J. Suppl. Ser.*, vol. 259, no. 2, pp. 64, Apr. 2022.

He Sun (Member, IEEE) received the bachelor's degree from Peking University, Beijing, China, in 2014, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 2019. He is currently an Assistant Professor with the College of Future Technology, Peking University. Prior to starting with Peking University, he was a Postdoctoral Researcher and Amazon AI4Science Fellow with Caltech, Pasadena, CA, USA. His research interests include computational imaging and adaptive optics, which tightly integrates optics, control, signal processing, and machine learning to design more capable scientific imaging instruments. His past work has contributed to interdisciplinary science missions, including the Event Horizon Telescope for black hole interferometric imaging and space telescope missions for exoplanet direct imaging.

Oscar Leong (Member, IEEE) received the undergraduation degree in mathematics from Swarthmore College, Swarthmore, PA, USA, in 2016, and the Ph.D. degree in computational and applied mathematics from Rice University, Houston, TX, USA, in 2021. He is currently a von Kármán Instructor with the Computing and Mathematical Sciences Department, California Institute of Technology, Pasadena, CA, USA. In July 2024, he will join the Statistics and Data Science Department, University of California, Los Angeles, Los Angeles, CA, USA, as a tenure-track Assistant Professor. His research interests include the mathematics of data science, inverse problems, and machine learning. He is broadly interested in using tools from convex geometry, high-dimensional statistics, and nonlinear optimization to better understand and improve data-driven inversion algorithms, along with their application to problems in the natural sciences. He was a NSF Graduate Fellow with Rice University. He is an MGB-SIAM Early Career Fellow and was selected as a Rising Star in Data Science (2022) by the University of Chicago, Chicago, IL, USA.

Katherine L. (Katie) Bouman received the bachelor's degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, and the Ph.D. degree in electrical engineering and computer science from the Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, USA. She is currently an Assistant Professor with the Computing and Mathematical Sciences, Electrical Engineering, and Astronomy Departments, California Institute of Technology, Pasadena, CA, USA. Her work combines ideas from signal processing, computer vision, machine learning, and physics to find and exploit hidden signals for scientific discovery. Before joining Caltech, she was a Postdoctoral Fellow with the Harvard-Smithsonian Center for Astrophysics. She is a Rosenberg Scholar, Heritage Medical Research Institute Investigator, recipient of the Royal Photographic Society Progress Medal, Electronic Imaging Scientist of the Year Award, University of Michigan Outstanding Recent Alumni Award, and co-recipient of the Breakthrough Prize in Fundamental Physics. As part of the Event Horizon Telescope Collaboration, she is co-lead of the Imaging Working Group and acted as coordinator for papers concerning the first imaging of the M87* and Sagittarius A* black holes.

Angela F. Gao received the bachelor's degree (with a double major) in electrical and computer engineering and biomedical engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2019. She is currently working toward the Ph.D. degree with the Computing and Mathematical Sciences Department, California Institute of Technology, Pasadena, CA, USA. Her research interests include imaging inverse problems, computational imaging, and generative models, with applications in scientific imaging problems.