Leveraging Common Structure for Prior-Free Image Reconstruction

Oscar Leong* **Angela F. Gao*** **He Sun** **Katherine L. Bouman**

Computing and Mathematical Sciences
California Insitute of Technology
{oleong, afgao, hesun, klbouman}@caltech.edu

Abstract

We consider solving ill-posed imaging inverse problems under a generic forward model. Due to the ill-posedness of such problems, prior models that encourage certain image-based structure are required to reduce the space of possible images when solving for a solution. Traditional approaches utilize hand-crafted prior models with parameters tuned through trial and error, which can be time-intensive and prone to human bias. Learning-based approaches use image samples from the ground-truth distribution of interest to learn a parameterized generative model that can be subsequently used to constrain the inverse problem; however, in many applications ground-truth images may be unavailable. In contrast, we propose to either select or learn an image generation model from the noisy measurements alone, without incorporating prior constraints on image structure. We first show how, given a number of candidate models, the Evidence Lower Bound (ELBO) of a variational distribution can be used to select an appropriate prior. Then, we showcase how, in the absence of available priors, we are able to directly learn an underlying generative model from a set of noisy measurements using a proxy for the ELBO. We crucially assume that the ground-truth images share a common structure by being drawn from the same underlying distribution. The learned model leverages this structure in its architecture, which consists of a shared generator with a compressed latent space, where each image posterior is learned variationally in the latent space. This allows the model to learn global properties of the image distribution without overfitting. We illustrate our framework on a variety of inverse problems, ranging from denoising to a compressed sensing problem inspired by black hole imaging.

1 Introduction

In imaging inverse problems, the goal is to recover a target clean image from corrupted measurements where the measurements and image are related via some understood forward model: $y = f(x) + \eta$. Here, y are our measurements, x is the ground-truth image, f is a forward model, and η is noise. Such problems are ubiquitous across the natural sciences, including denoising [6], super-resolution [7], compressed sensing [8, 10, 11], phase retrieval [12], and deconvolution [21]. When the problem is ill-posed, there are many images that are consistent with our measurements. Thus, we require structural assumptions, referred to as image priors, to reduce the space of possible solutions. In this paper, we parameterize prior image distributions using an *image generation model (IGM)*.

In order to define a prior via an IGM, it is necessary to have knowledge of the structure of the underlying image distribution. However, there are many scientific imaging modalities (e.g., medical

*equal contribution

Preprint. Under review.

imaging, geophysical imaging, and astronomical imaging) where we do not have access to ground-truth images. Collecting ground-truth images in these domains can be extremely invasive, time-consuming, expensive, or even impossible. Moreover, while classical approaches utilize hand-crafted image models [15, 23] to regularize inversion, hyperparameter tuning is imperative to success. In practice, this results in choosing a regularizer based on trial and error and such approaches are heavily prone to human bias [20]. If ground-truth images are available, then an IGM can be learned directly [28, 34, 4], but this approach requires access to an abundance of clean data. Thus, developing a principled *criterion* to select an IGM based on access to noisy measurements alone could significantly reduce modelling time and avoid any imposed human bias. If such a criterion can be identified, a natural follow-up question would be to use such a criterion to *learn* the underlying IGM from noisy corrupted measurements alone.

In this work, we solve a collection of ill-posed image reconstruction tasks without access to an explicit image prior. The key insight of our work is that knowledge of common structure across independent problems is sufficient regularization alone. In particular, we first show how, given a number of candidate IGMs, the Evidence Lower Bound (ELBO) of a variational distribution can be used to select an image model that best explains the underlying true image. The motivation for our approach lies in the fact that given measurements y and an IGM m, the ELBO provides a tractable proxy for the IGM posterior distribution $p(m|y)$. Then, we show how one can, in the absence of available models, directly learn an IGM from a set of N noisy measurements $y^{(i)} = f(x^{(i)}) + \eta^{(i)}$ to help solve the underlying inverse problem $y \mapsto x$. An important assumption we make is that the underlying images $x^{(i)}$ are drawn from the same distribution (unknown a-priori), and we show how the learned model m is able to capture this common structure. Our contributions are as follows:

1. We mathematically and experimentally illustrate how a proxy for the ELBO provides a good criterion for selecting an IGM in several inverse problems.
2. We solve inverse problems without a pre-defined IGM by directly learning it from an independent set of corrupted measurements. In particular, we parameterize the IGM as a deep generative network whose latent space is shared across measurements in order to leverage similarities in the underlying truth images.

1.1 Background and Related Work

Model selection. Model selection techniques seek to choose models that best explain 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 [31]. However, in image reconstruction or other inverse problems with limited data, one cannot afford to hold out data. In these cases, a more common way is to conduct model selection based on probabilistic metrics. The simplest probabilistic metric used for linear model selection is adjusted R^2 [24]. It re-weights the goodness-of-fit by the number of linear model parameters, so it helps reject high dimensional parameters that do not improve the data fitting accuracy. Similar metrics in non-linear model selection are Bayesian Information Criterion (BIC) [29] 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.

Learning IGMs. There are a few recent methods that learn IGMs from only noisy data. Regularization by Artifact Removal (RARE) [22] learns data-driven priors from various forms of "artifacts" (e.g. compressed sensing aliasing). RARE's Artifact2Artifact scheme requires multiple independent measurements of each image in a collection of images to train an IGM. However, this approach is limited by the assumption that the expected value of multiple observations of a single image is the ground-truth image. Techniques [5, 16] based on Generative Adversarial Networks (GANs) have also been proposed to learn an IGM directly from noisy data. For instance, AmbientGAN [5] learns to generate images whose measurements are indistinguishable from the true measurements. Although the GAN based techniques can achieve impressive results, as we show in Section 3, they require many independent observations to achieve this high performance.

2 Approach

In this work, we propose to either select or learn an IGM from a collection of corrupted measurements with underlying images that share a common image distribution. In particular, we first consider selecting an IGM by choosing the model that approximately maximizes the model posterior by using the ELBO as a principled proxy. Building upon this, we show that one can directly learn the IGM from noisy measurements alone by parameterizing the IGM as a deep neural network with a low-dimensional latent distribution. The IGM network weights are shared across latent posterior distributions, capitalizing on the common structure shared across images.

2.1 Motivation for ELBO as a model selection criterion

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 a prior, or IGM m, that best explains our data. A natural approach would be to find or select the model m that maximizes the model posterior distribution $p(m|y) \propto p(y|m)p(m)$. That is, conditioned on the noisy measurements, find the model of highest likelihood. Unfortunately computing $p(y|m)$ is intractable, but we show how it can be well approximated using the Evidence Lower Bound (ELBO).

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

$$\begin{aligned} D_{\mathrm{KL}}(q_\theta(x) \parallel p(x|y,m)) &:= \mathbb{E}_{x\sim q_\theta(x)}[\log q_\theta(x) - \log p(x|y,m)] \\ &= \log p(y|m) - \mathbb{E}_{x\sim q_\theta(x)}\left[\log p(y|x,m) + \log p(x|m) - \log q_\theta(x)\right] \end{aligned}$$

The ELBO of a model m given measurements y is defined by

$$\mathrm{ELBO}(m; y) := \mathbb{E}_{x\sim q_\theta(x)}\left[\log p(y|x,m) + \log p(x|m) - \log q_\theta(x)\right]. \tag{1}$$

Rearranging the above equation, we see that

$$\log p(y|m) = D_{\mathrm{KL}}(q_\theta(x) \parallel p(x|y,m)) + \mathrm{ELBO}(m; y) \geqslant \mathrm{ELBO}(m; y)$$

by the non-negativity of the KL-divergence. This leads to a criterion to choose a model:

$$\log p(m|y) \geqslant \mathrm{ELBO}(m; y) + \log p(m).$$

Note that if the variational distribution $q_\theta(x)$ is a good approximation to the posterior $p(x|y, m)$, $D_{\mathrm{KL}} \approx 0$ and $\log p(m|y) \approx \mathrm{ELBO}(m; y) + \log p(m)$.

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

ELBOProxy. Some common image priors are explicit, which allows for direct computation of $\log p(x|m)$. In this case, we can optimize the ELBO defined in Equation 1 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 preventing overfitting to noise. However, this means that we can only compute $\log p(x|m)$ directly if we have an injective map [17]. This architectural requirement limits the expressivity of the network. Hence, we instead consider a proxy of the ELBO that is especially helpful for deep generative networks. Suppose our image generation model is of the form $x = G_m(z)$ where G_m is a generative network and z is a latent vector. Introducing a variational family for our latent representations $z \sim q_\phi(z)$ and using $\log p(z|G_m)$ in place of $\log p(x|G_m)$, we arrive at the following proxy of the ELBO:

$$\mathrm{ELBOProxy}(G_m; y) := \mathbb{E}_{z\sim q_\phi(z)}\left[\log p(y|G_m(z)) + \log p(z|G_m) - \log q_\phi(z)\right]. \tag{2}$$

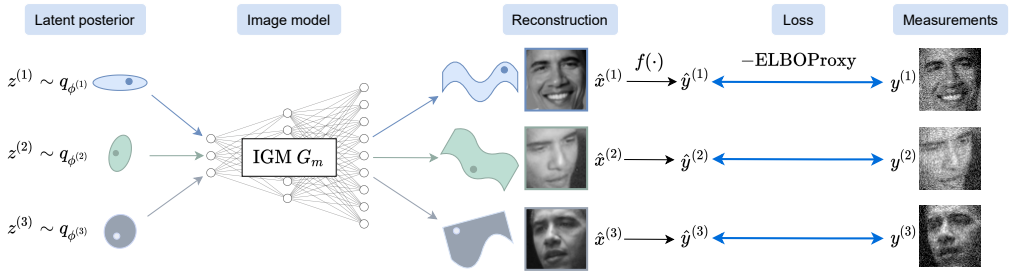

Figure 1: We propose to learn a single IGM G_m from a set of noisy measurements $\{y^{(i)}\}_{i=1}^N$ to reconstruct each underlying image $x^{(i)}$. This generator is shared across latent posteriors and learns global properties of the underlying image distribution. We reconstruct each image by learning a reconstruction posterior over the latent space of G_m, each parameterized by a variational family $q_{\phi^{(i)}}$ with parameters $\phi^{(i)}$. The IGM parameters m and parameters $\phi^{(i)}$ are found by solving Equation 3.

Variational family. In practice, there are a number of choices to parameterize the variational families $q_\theta(x)$ and $q_\phi(z)$. One could, for example, utilize a Gaussian parameterization which would entail learning a mean and covariance (μ, Σ). Another particularly flexible family of functions are Normalizing Flows [27, 26, 9], which are generative models capable of learning an invertible mapping between a simple latent distribution to a more complicated distribution of interest [32]. We explore both options in this work, and discuss these ideas further in the subsequent sections.

Toy example. In order 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 plausible image models. For this experiment, we use the MNIST dataset [19] and consider two inverse problems: denoising and phase retrieval. We train a generative model G_{m_c} on each class $c \in \{0, 1, 2, \ldots, 9\}$. Hence, G_{m_c} is learned to generate images from class c via $G_{m_c}(z)$ where $z \sim \mathcal{N}(0, I)$. Then, given noisy measurements y_c from an image from class c, we ask whether the generative model G_{m_c} from the appropriate class would achieve the best ELBOProxy. 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 $-$ ELBOProxy obtained by using model $G_{m_{j-1}}$ to reconstruct images from class $i - 1$. We calculate $\text{ELBOProxy}(G_{m_c}; y_c)$ by parameterizing q_{ϕ_c} with a Normalizing Flow and optimizing network weights ϕ_c. Results from the first 5 classes are shown in Fig. 2 and the full arrays are shown in the Appendix. We note that all of the correct models are chosen in phase retrieval and 3 out of 5 models are chosen correctly in denoising, although the difference in value is small between the true class and the incorrectly selected class. 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_{m_4} and G_{m_9} achieve comparable ELBOProxy values. By carefully inspecting the noisy image of the 4 one can see that both models are reasonable given the structure of the noise.

2.2 Learning the image generation model for inverse problems

As the previous section illustrates, the ELBOProxy provides a good criterion for choosing an appropriate IGM from noisy measurements. However, there are a number of challenging imaging inverse problems where we do not have access to a set of potential IGMs that contain an accurate image model. Here, we consider the task of learning the IGM directly from the noisy data. We consider the setting where we have access to a collection of N measurements $y^{(i)} = f(x^{(i)}) + \eta^{(i)}$ for $i \in [N]$ and each $x^{(i)}$ are drawn from the same image distribution of interest. For example, we may be imaging the same object from various viewpoints, but our observations are corrupted and noisy. The key assumption we make is that similar structure exists in all underlying images measured, allowing us to learn the structure by finding a low-dimensional representation that can produce all noisy data observed. Relative to typical generative modelling-based approaches or supervised learning approaches, we assume we have very few examples N, on the order of only 10's of examples.

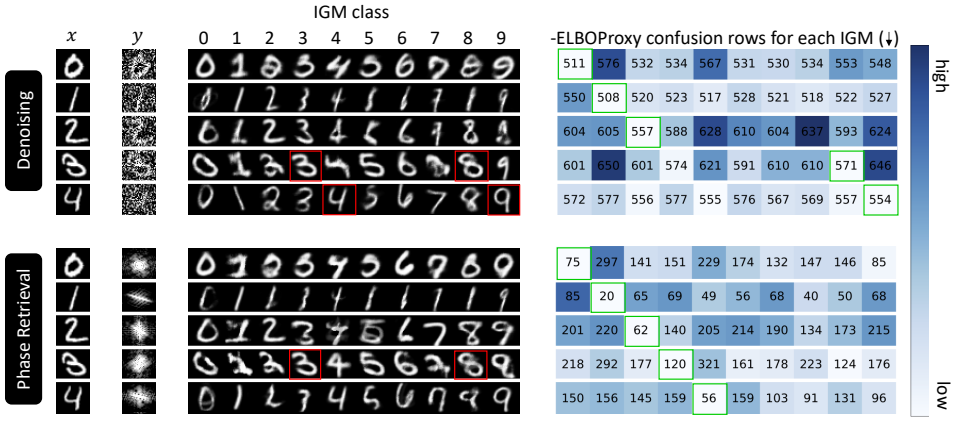

Figure 2: We consider two inverse problems: denoising and phase retrieval. Left: the two leftmost columns correspond to the ground truth image x_c and the noisy measurements y_c. Center: we show the means of the variational distributions given by the IGM trained in a particular class. Interesting examples from a model selection perspective are highlighted in red. Right: each row of the array corresponds to the $-$ ELBOProxy achieved by each model in reconstructing the images. Here, lower is better. Boxes highlighted in green correspond to the best $-$ ELBOProxy values in each row.

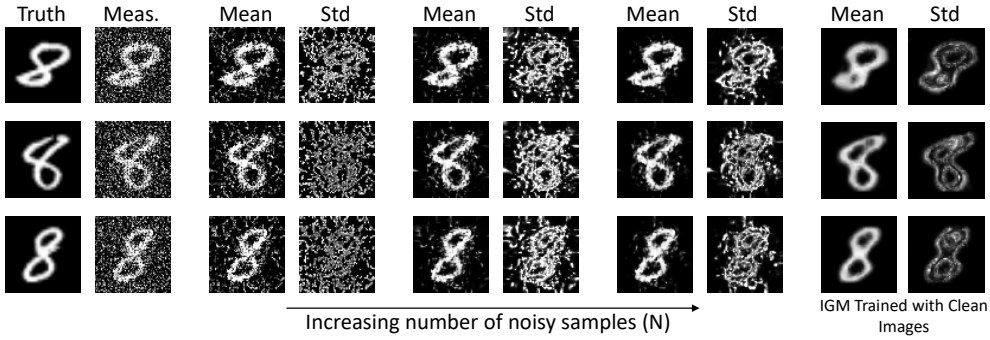

Figure 3: **Denoising improves with more noisy MNIST observations.** We demonstrate our method of learning the IGM to perform denoising for increasing number of noisy images (4, 35, and 75 images from left to right). In each panel, we include the ground truth, noisy measurements, mean of the posterior, and standard deviation of the posterior. We also include the reconstructions using an IGM trained on the full clean MNIST 8's class. 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.

Learning approach. We would like our image generation model to capture shared properties of the N images underlying the set of measurements. Each corruption, however, could induce its own complicated image posteriors. Thus, we propose to find a *shared* generator G_m with weights m that can be used to reconstruct the full posterior of each image $x^{(i)}$ from its corresponding noisy observation $y^{(i)}$. This approach is illustrated in Fig. 1. More explicitly, given a set of noisy measurements $\{y^{(i)}\}_{i=1}^N$, we optimize the ELBOProxy from Equation 2 to learn a generator G_m that leads to an accurate set of variational distributions $\{q_{\phi^{(i)}}\}_{i=1}^N$:

$$\max_{m,\{\phi^{(i)}\}} \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{z\sim q_{\phi^{(i)}}(z)}\left[\log p(y^{(i)}|G_m(z)) + \log p(z|G_m) - \log q_{\phi^{(i)}}(z)\right] + \log p(G_m). \quad (3)$$

The expectation in this objective is approximated via Monte Carlo sampling. In terms of choices for $\log p(G_m)$, we can add additional regularization to promote particular structures, such as smoothness. Here, we consider having sparse neural network weights as a form of regularization and use dropout during training to represent $\log p(G_m)$[30].

Once a generator G_m and variational parameters $\phi^{(i)}$ have been learned, we can solve the inverse problem by providing estimates of $x^{(i)}$ via sampling $\hat{z} \sim q_{\phi^{(i)}}(z^{(i)})$ and outputting $G_m(\hat{z}) \approx x^{(i)}$ or computing an average $\frac{1}{T}\sum_{t=1}^{T} G_m(\hat{z}_t) \approx x^{(i)}$. Producing samples can help visualize the range of uncertainty under the current image model G_m, 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.

Toy Example. To illustrate learning the IGM, we consider using the ELBOProxy to learn the IGM directly from a set of noisy images. The noisy images $\{y^{(i)}\}_{i=1}^{N}$ are defined by $y^{(i)} = x^{(i)} + \eta^{(i)}$ where $\{x^{(i)}\}_{i=1}^{N}$ are from a single MNIST class c and $\eta^{(i)} \sim N(0, \sigma^2 I)$ where $\sigma = 0.5$. Using only these noisy images, we learn the shared generator G_m as well as unique variational posterior distributions that each represents a single denoised reconstruction. We use multivariate Gaussian distributions to parameterize the variational posteriors and a Deep Decoder [13] as the generator with dropout of 10^{-4}. Results from this experiment are shown in Fig. 3. Note that as the number of independent noisy measurements increases, the IGM becomes a stronger prior that better captures the true underlying distribution, resulting in higher fidelity reconstructions. In addition, as the number of images increases, the uncertainty focuses on the salient features of the image rather than on the noise.

3 Results

We now consider solving various inverse problems by learning an IGM directly from noisy measurements via the framework described in 2.2. For each of these experiments, we use a multivariate Gaussian distribution to represent each of the posterior distributions $z^{(i)} \sim q_{\phi^{(i)}}(z^{(i)})$ and a Deep Decoder as the IGM with a dropout of 10^{-4} for the rest of the experiments. The multivariate Gaussian distribution is parameterized by means $\{\mu^{(i)}\}_{i=1}^{N}$ and covariance matrices $\{\Lambda^{(i)} = L^{(i)}L^{(i)T} + \varepsilon I\}_{i=1}^{N}$, where εI with $\varepsilon = 10^{-3}$ is added to the covariance matrix to help with stability of the optimization. We choose to parameterize the posterior using simple Gaussian distributions due to memory constraints. Note that we do not employ early stopping. For more details, please see the Appendix.

Figure 4: **Denoising 75 images of a celebrity.** We demonstrate our method described in 2.2 using 75 noisy images of a celebrity. Here we show the ground truth (top), noisy measurements (middle), and mean reconstruction (bottom) for a subset of the 75 different noisy images. Our reconstructions are much less noisy, exhibiting sharper features that are hard to discern in the noisy images. Note that no predefined prior/regularizer was used in denoising.

Denoising: We show results on denoising noisy images of 8's from the MNIST dataset in Fig 3 and denoising noisy images from a single face from the PubFig [18] dataset in Fig. 4. The measurements for both are defined by $y = x + \eta$ where $\eta \sim \mathcal{N}(0, \sigma^2 I)$ with an SNR of $\sim$0.5 for the MNIST digits and an SNR of $\sim$54 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. Note that the learned IGM improves as the number of independent observations increases, as shown in Fig. 3. Our reconstructions also substantially outperform the baseline methods AmbientGAN [5], Deep Image Prior (DIP) [33], and regularized maximum likelihood using total variation (TV-RML), as shown in Fig. 7. Unlike DIP, our method does not seem to overfit and does not require early stopping. Our method does not exhibit noisy artifacts like those seen in AmbientGAN results.

Compressed sensing: We consider noisy measurements from the following problem arising from astronomical imaging of black holes with the Event Horizon Telescope (EHT): suppose we are given

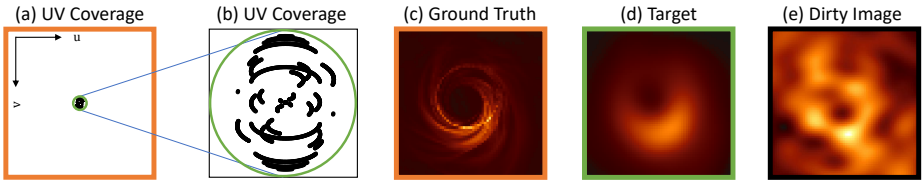

Figure 5: **Visualization of the intrinsic resolution of the EHT compressed sensing measurements.** The EHT measures sparse spatial frequencies of the image (i.e., components of the image's Fouier transform). In order to generate the ground truth image (c), all frequencies in the entire domain of (a) must be used. Restricting spatial frequencies to the ones in (a) and (b)'s green circle generates the target (d). The EHT samples a subset of this region, 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 $\hat{x} = A^H y$. The 2D spatial Fourier frequency coverage represented with (u, v) positions is referred to as the UV coverage.

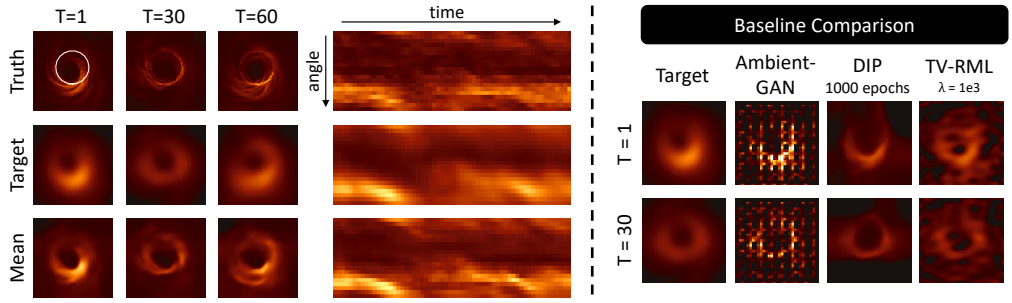

Figure 6: **Compressed sensing with a video of a black hole.** We demonstrate our method described in 2.2 using 60 images from an evolving black hole target. Left: Here we show the ground-truth, target, and mean reconstruction, respectively. Additionally, we show the unwrapped space × time image, which is taken along the white ring illustrated in the T=1 ground-truth image. The bright-spot's temporal trajectory of our reconstruction matches that of the truth and target. 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.

access to measurements of the form $y = Ax + \eta,\ \eta \sim \mathcal{N}(0, \sigma^2 I)$ where $A \in \mathbb{C}^{p \times n}$ is a low-rank compressed sensing matrix arising from interferometric telescope measurements. This problem is highly ill-posed and requires the use of priors or regularizers to recover a reasonable image [2]. Moreover, it is impossible to acquire ground-truth images of black holes, so any hand-designed prior will exhibit human bias. However, we can assume that the source does not change much day to day, which matches the assumptions of our approach and motivates the use of an IGM.

We show results on 60 frames from a video of a simulated evolving black hole target [25, 3] with an SNR of $\sim$7 in Fig. 6. Our reference target image is the ground-truth filtered with a low pass filter that represents the maximum resolution intrinsic to the Event Horizon Telescope array, which is visualized and explained in Fig. 5. Our method is not only able to reconstruct the large scale features of the ground-truth image without any aliasing artifacts, but also achieve some level of super-resolution. Our reconstructions also achieve higher super-resolution as compared to our baselines in Fig. 6 and do not exhibit artifacts evident in the AmbientGAN and TV-RML baselines. We additionally note that the reconstructions from DIP are strong in this case, which is partially due to the fact that such networks are not prone to overfitting in compressed sensing problems [14].

Phase retrieval: Here we demonstrate a limitation of our approach with an example where our method obtains sub-optimal performance. In particular, we consider the non-linear inverse problem of phase retrieval. Our measurements are described by $y^{(i)} = |\mathcal{F}(x^{(i)})| + \eta^{(i)}$ where $\mathcal{F}$ is the Fourier transform and $\eta^{(i)} \sim \mathcal{N}(0, \sigma^2 I)$. Since each measurement is the magnitude of the Fourier transform, possible reconstructed images include all spatial shifts. Due to the severe ill-posedness of the problem, representing this complicated posterior that includes all spatial shifts is challenging.

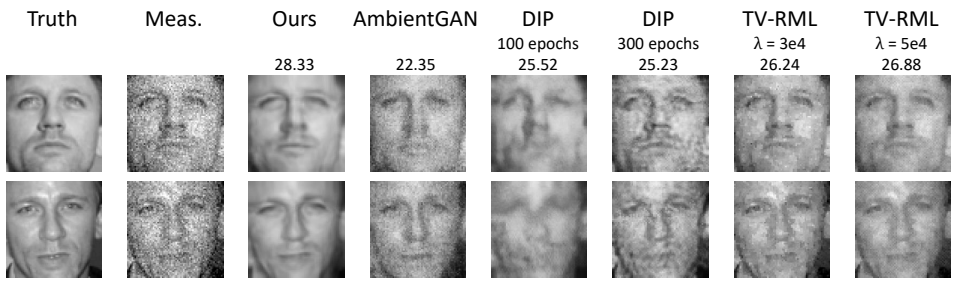

Figure 7: **Denoising baseline comparisons.** We compare to various baselines (AmbientGAN, Deep Image Prior (DIP), and regularized maximum likelihood using TV (TV-RML) with weight λ, and we report the average PSNR across all 75 reconstructions. We show both early stopping and full training results using DIP. Our method exhibits substantially higher PSNR and less noise than all other baselines while maintaining distinct features that are smoothed out by DIP.

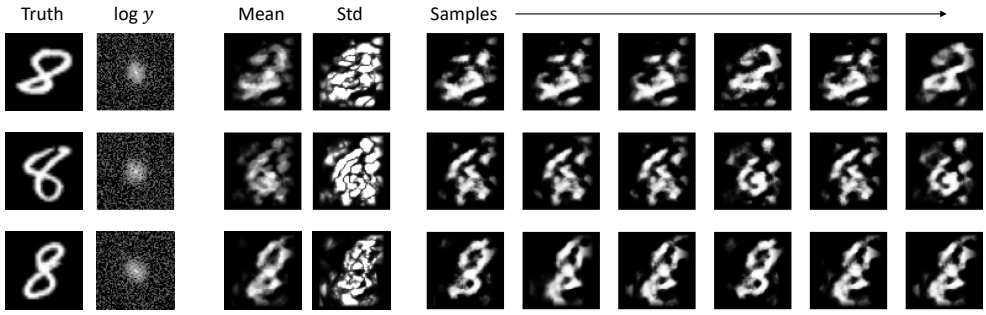

Figure 8: **Phase retrieval from MNIST 8's.** We demonstrate our method described in 2.2 to perform phase retrieval on 75 images. Here we show the target image, the log magnitude of the Fourier transform, mean of the posterior, standard deviation of the posterior, and samples from the posterior. Note that the posterior samples exhibit diversity even with a simple Gaussian latent distribution.

Thus, we incorporate an envelope as the final layer of G_m to encourage the reconstruction to be centered. Nonetheless, multiple shifts are still possible within this enveloped region.

We show results from 75 noisy phase retrieval measurements from the MNIST 8's class with an SNR of $\sim$0.04 in Fig. 8. Our reconstructions have features similar to the digit 8, but contain significant artifacts. These artifacts are due to the fact that a highly multi-modal distribution is being represented by a unimodal Gaussian variational distribution. In denoising, increasing the number of measurements N improved the IGM; however, this shift ambiguity will remain even if we increase the number of available measurements. This problem would benefit from a translation-invariant generator G_m, which would simplify the complexity of the posterior distribution. We save this for future work.

3.1 Generalization to new data and forward models

We consider generalizing to new measurements given an IGM G_m learned using the methods described in Section 2.2. These measurements could be from the same forward model (e.g., the IGM was trained with noisy MNIST digits, and we want to use it to denoise an unseen MNIST digit) or from a different forward model (e.g., the IGM was trained with noisy MNIST digits, and we want to use it to perform compressed sensing on an MNIST digit), but the measurements are all from the same underlying ground-truth image distribution that the IGM was originally trained with. To solve the new inverse problem, we learn the latent posterior $z \sim q_\phi(z)$ given fixed weights m of the IGM G_m and our measurement y using Eq. 2. Due to the additional complexity in generalizing to a new image and forward model, we parameterize q_ϕ with a Normalizing Flow model. We show results in Fig. 9 showcasing the generalization performance of an IGM pre-trained on 75 noisy images to 1) novel images under the same forward model and 2) novel images under a different forward model. Our method is able to reconstruct primary features of the face in both generalization tests. This

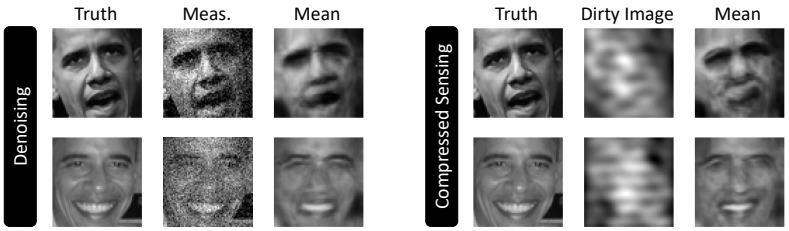

Figure 9: **Generalization to novel measurements.** We show results using an IGM trained on 75 noisy images to solve inverse problems on novel measurements/images. These ground-truth images were not used during training. We perform denoising on the left and compressed sensing on the right. Note that when performing compressed sensing, there are two sources of novelty: the underlying ground-truth image and the forward model.

demonstrates that the IGM avoids overfitting to specific images and measurements, and is able to learn generalizable properties of the underlying data distribution even from few examples.

4 Conclusion

In this work, we demonstrated how one can solve ill-posed image reconstruction problems without pre-defined priors by jointly learning a shared low-dimensional structure of the ground-truth images from corrupted measurements alone. We first demonstrated that, given a number of candidate IGMs, the ELBO of a variational distribution can be used to select the best model. By leveraging common structure present across the underlying ground-truth images, we then showed that one can directly learn an IGM from a set of corrupted measurements by searching for models that maximize a proxy of the ELBO. We demonstrated that our approach learns global properties of the underlying image distribution and can successfully solve a diverse set of inverse problems with significantly fewer examples than generative model-based approaches. Overall, our work showcases the possibilities of solving inverse problems in a "prior-free" fashion, free from human bias traditionally encountered in ill-posed image reconstruction. Moreover, we believe our approach could aid in automatically discovering novel structure from scientific measurements without access to clean data, leading to potentially new avenues for scientific discovery.

Broader Impacts. We show that the ELBO can be used to perform model selection. However, this does not guarantee that the selected model is a good enough IGM for accurate inverse image reconstruction. Nonetheless, our approach can improve our understanding of fundamental science to see things that were previously invisible.

Acknowledgements and Disclosure of Funding. We declare no competing interests.

References

[1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723, 1974. (Cited on 2)

[2] Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *The Astrophysical Journal Letters*, 875(1):L4, 2019. (Cited on 7)

[3] Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, David Ball, Mislav Baloković, John Barrett, Dan Bintley, et al. First m87 event horizon telescope results. v. physical origin of the asymmetric ring. *The Astrophysical Journal Letters*, 875(1):L5, 2019. (Cited on 7)

[4] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros Dimakis. Compressed sensing using generative models. *International Conference on Machine Learning (ICML)*, 2017. (Cited on 2)

[5] Ashish Bora, Eric Price, and Alexandros G Dimakis. Ambientgan: Generative models from lossy measurements. In *International conference on learning representations*, 2018. (Cited on 2, 6)

[6] Harold C. Burger, Christian J. Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. (Cited on 1)

[7] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2013. (Cited on 1)

[8] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. (Cited on 1)

[9] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. 2017. (Cited on 4)

[10] David Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6), 2006. (Cited on 1)

[11] David Donoho, Michael Lustig, and John M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. (Cited on 1)

[12] Albert Fannjiang and Thomas Strohmer. The numerics of phase retrieval. *Acta Numerica*, 29:125 – 228, 2020. (Cited on 1)

[13] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *International Conference on Learning Representations (ICLR)*, 2019. (Cited on 6)

[14] Reinhard Heckel and Mahdi Soltanolkotabi. Compressive sensing with un-trained neural networks: Gradient descent finds the smoothest approximation. *International Conference on Machine Learning*, 2020. (Cited on 7)

[15] Leonid I.Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60:259–268, 1992. (Cited on 2)

[16] Maya Kabkab, Pouya Samangouei, and Rama Chellappa. Task-aware compressed sensing with generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. (Cited on 2)

[17] Konik Kothari, AmirEhsan Khorashadizadeh, Maarten de Hoop, and Ivan Dokmanić. Trumpets: Injective flows for inference and inverse problems. In *Uncertainty in Artificial Intelligence*, pages 1269–1278. PMLR, 2021. (Cited on 3)

[18] Neeraj Kumar, Alexander C Berg, Peter N Belhumeur, and Shree K Nayar. Attribute and simile classifiers for face verification. In *2009 IEEE 12th international conference on computer vision*, pages 365–372. IEEE, 2009. (Cited on 6)

[19] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278—-2324, 1998. (Cited on 4)

[20] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1964–1971, 2009. (Cited on 2)

[21] Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. Rapid, robust, and reliable blind deconvolution via nonconvex optimization. *arXiv preprint*, arXiv:1606.04933, 2016. (Cited on 1)

[22] Jiaming Liu, Yu Sun, Cihat Eldeniz, Weijie Gan, Hongyu An, and Ulugbek S Kamilov. Rare: Image reconstruction using deep priors learned without groundtruth. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1088–1099, 2020. (Cited on 2)

[23] Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999. (Cited on 2)

[24] Jeremy Miles. R squared, adjusted r squared. *Encyclopedia of Statistics in Behavioral Science*, 2005. (Cited on 2)

[25] Oliver Porth, Koushik Chatterjee, Ramesh Narayan, Charles F Gammie, Yosuke Mizuno, Peter Anninos, John G Baker, Matteo Bugli, Chi-kwan Chan, Jordy Davelaar, et al. The event horizon general relativistic magnetohydrodynamic code comparison project. *The Astrophysical Journal Supplement Series*, 243(2):26, 2019. (Cited on 7)

[26] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *International Conference on Machine Learning (ICML)*, 2015. (Cited on 4)

[27] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. *International Conference on Machine Learning (ICML)*, 4:3057—3070, 2014. (Cited on 4)

[28] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. (Cited on 2)

[29] Gideon E. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978. (Cited on 2)

[30] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. (Cited on 5)

[31] Mervyn Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, 36(2):111–133, 1974. (Cited on 2)

[32] He Sun and Katherine L. Bouman. Deep probabilistic imaging: Uncertainty quantification and multi-modal solution characterization for computational imaging. *35th AAAI Conference on Artificial Intelligence*, 2021. (Cited on 4)

[33] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018. (Cited on 6)

[34] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE, 2013. (Cited on 2)