

---

# Variational Inference for Computational Imaging Inverse Problems

---

Francesco Tonolini<sup>1</sup> Ashley Lyons<sup>2</sup> Piergiorgio Caramazza<sup>2</sup> Daniele Faccio<sup>2</sup> Roderick Murray-Smith<sup>1</sup>

## Abstract

We introduce a method to infer a variational approximation to the posterior distribution of solutions in computational imaging inverse problems. Machine learning methods applied to computational imaging have proven very successful, but have so far largely focused on retrieving a single optimal solution for a given task. Such retrieval is arguably an incomplete description of the solution space, as in ill-posed inverse problems there may be many similarly likely reconstructions. We minimise an upper bound on the divergence between our approximate distribution and the true intractable posterior, thereby obtaining a probabilistic description of the solution space in imaging inverse problems with empirical prior. We demonstrate the advantage of our technique in quantitative simulations with the CelebA dataset and common image reconstruction tasks. We then apply our method to two of the currently most challenging problems in experimental optics: imaging through highly scattering media and imaging through multi-modal optical fibres. In both settings we report state of the art reconstructions, while providing new capabilities, such as estimation of error-bars and visualisation of multiple likely reconstructions.

## 1. Introduction

In recent years, machine learning has taken an increasingly central role in solving computational imaging (CI) problems, leading to unprecedented reconstruction quality in a range of settings (Lucas et al., 2018; Kulkarni et al., 2016; Chang et al., 2017) and allowing for new and challenging imaging tasks to be accomplished (Liu et al., 2015a; Isola et al., 2017). However, learning methods have so far been applied to inverse imaging problems to infer a single near-optimal

reconstruction, lacking a probabilistic explanation of the solution space. Conversely, as advancements in machine learning allow for increasingly ill-posed imaging inverse problems to be approached, probabilistic descriptions become important; if appreciably different solutions satisfy similarly well the observed measurements and empirical prior, an accurate retrieval should capture their variability.

In this paper, we introduce a variational method for approximating the posterior distribution in ill-posed imaging inverse problems, where the prior is empirically learned from signal examples. Given some model of the forward observation process and a set of signal examples, we minimise stochastically an upper bound on the cross entropy between the true posterior and a flexible parametric distribution. With the new method, we obtain better reconstruction performance than inference using equivalent deterministic neural networks, while at the same time retrieving information which captures the nature of the posterior density, such as accurate pixel-wise error-bars and draws from the distribution of likely solutions.

We first demonstrate the advantage of our method in quantitative simulations, retrieving images of faces from the CelebA dataset (Liu et al., 2015b) under different common degradation conditions. We then perform reconstructions from experimental acquisitions in two particularly ill-posed imaging settings, currently at the focus of optical imaging research: computational diffuse imaging (Badon et al., 2016; Durán et al., 2015; Boccolini et al., 2017; Horisaki et al., 2016) and imaging through multi-modal optical fibres (Plöschner et al., 2015; Turtaev et al., 2018; Moran et al., 2018). In both sets of experiments, we report state of the art reconstructions, while clearly demonstrating the importance of capturing the nature of variation in the solution space through our variational approach.

## 2. Background and Related Work

Computational imaging (CI) broadly identifies a class of methods in which an image of interest is not directly observed, but rather inferred algorithmically from one or more measurements (Bertero & Boccacci, 1998). Many image recovery tasks fall within this definition, such as deconvolution (Starck et al., 2003; Bioucas-Dias et al., 2006), computed tomography (CT) (Ritschl et al., 2011) and struc-

<sup>1</sup>School of Computing Science, University of Glasgow, UK

<sup>2</sup>School of Physics and Astronomy, University of Glasgow, UK..  
Correspondence to: 2402432t <@student.gla.ac.uk>.

We acknowledge funding from Amazon and EPSRC grants EP/M01326X/1, EP/R018634/1.

tured illumination imaging (Magalhães et al., 2011).

Traditionally, CI tasks are modelled as inverse problems, where a target signal  $x \in \mathbb{R}^n$  is measured through a forward model  $y = f(x)$ , yielding observations  $y \in \mathbb{R}^m$ . The aim is then to retrieve the signal  $x$  from the observations  $y$  (Bertero & Boccacci, 1998; Vogel, 2002). In the following subsections, we review the main advances in solving imaging inverse problems, starting from linear models and user defined regularisation functions, to then focus on the more recent application of learning based methods.

### 2.1. Linear models and Hand-crafted Priors

In many CI problems, the forward model can be described as a linear operator  $A \in \mathbb{R}^{m \times n}$  and some independent noise  $\epsilon \in \mathbb{R}^m$ , such that  $f(x) = Ax + \epsilon$  (Bertero & Boccacci, 1998; Daubechies et al., 2004). Under such an assumption, the aim is to find a solution image which satisfies the linear observations well, while retaining some properties of interest. The estimate of  $x$  is then recovered by minimising an objective function of the form

$$\frac{1}{2} \|Ax - y\|^2 + \lambda h(x), \quad (1)$$

where  $\|\cdot\|$  indicates the Euclidean norm and  $h(x)$  is a penalty function that enforces some desired property in  $x$ . In the case of images, it is common to assume that  $x$  is sparse in some basis, such as frequency or wavelets, leading to  $\ell_1$ -norm penalty functions of the form  $h(x) = \|Wx\|_1$ , where  $W$  is a unitary operator which maps  $x$  to a sparse basis (Daubechies et al., 2004; Donoho, 2006; Figueiredo et al., 2007). For such choices of penalty, the objective function of equation 1 is convex. This makes the optimisation problem solvable with a variety of efficient methods (Daubechies et al., 2004; Yang & Zhang, 2011) and provides theoretical guarantees on the recoverability of the solution (Candes et al., 2006).

The aforementioned framework has been widely applied to perform CI. For instance, many image restoration tasks, such as de-blurring, up-sampling and in-painting have been formulated as ill-conditioned linear inverse problems and are solved as described above (Osher et al., 2005). Various more complex sensing models can also be cast as linear operators, leading to the use of constrained optimisation in several CI systems that rely on ill-posed observations, such as sparse CT (Ritschl et al., 2011), single pixel photography (Sun et al., 2013) and imaging of objects hidden from view (Velten et al., 2012).

However, despite its theoretical benefits and wide adoption, sparsity regularisation is often too generic to accurately model the desired assumptions in a given imaging setting (Chang et al., 2017).

### 2.2. Bayesian Inference in Linear Inverse Problems

Bayesian inference has been applied to approximate the posterior distribution in linear inverse problems such as those described in the previous section, where the prior distribution is defined by some assumed image property, such as sparsity in some basis or image smoothness (Ji et al., 2008; Mohammad-Djafari, 2013). In particular, variational inference has been applied to common inverse problems such as blind deconvolution and image super-resolution (Likas & Galatsanos, 2004; Babacan et al., 2011). In these works, the true intractable posterior of the inverse problem is approximated with a tractable parametric function.

Differently from these methods, we aim to use examples of images as signal prior, rather than defined functional assumptions. This makes the techniques described above not directly applicable to our target case, as the prior can not be point evaluated, but we only have access to a finite number of draws from it in the form of a data set of signal examples.

### 2.3. Learning Inverse Mappings

The increasing availability of image data sets and continuous advancements in learning inference models enabled new possibilities for CI and inverse problems in general. Most learning approaches can arguably be described as inverse mapping models; with enough example pairs available, a neural network can be trained to directly recover a signal  $x$  from observations  $y$  (Lucas et al., 2018).

With such approaches, because a direct inverse mapping is learned, there are a number of advantages compared to optimisation with sparse regularisation. First, the model is trained with a set of images the target solution is assumed to belong to, implicitly making the signal assumptions more specific than sparsity in some basis. Second, the observation model  $f(x)$  is not constrained to be linear, or even differentiable; so long as we have access to a large number of signal-observations pairs we can train the network to perform the inversion. Third, once the model is trained, inference is non-iterative and thus typically much faster.

Several methods have been developed to perform such inference in a variety of CI settings (Egmont-Petersen et al., 2002; McCann et al., 2017). State of the art performance has been demonstrated with specifically designed neural networks models in many common image processing tasks, such as deconvolution and super-resolution (Xu et al., 2014; Ledig et al., 2017), as well as signal recovery from under determined projective measurements (Kulkarni et al., 2016).

Despite their proven ability to recover compelling images in CI inverse problems, empirical inverse mappings often lack model consistency, as recovered solutions are not directly induced to satisfy the observed measurements, but only de-

terminated by a flexible approximator trained with arbitrarily many signal-observations example pairs.

## 2.4. MAP Inference with Generative Models

A second emerging class of learning based methods in CI is that of maximum *a posteriori* (MAP) inference with generative models. In these frameworks, similarly to MAP inference with analytical priors, the solution is found by minimising the Euclidean distance to the observations  $\|f(x) - y\|$ . However, instead of regularising the solution with some penalty function, the signal of interest is assumed to be generated by a pre-trained generative model as  $x = g(z)$ , where  $z \in \mathbb{R}^k$  is the generative model’s latent variable (Bora et al., 2017). As the generative model is trained with an example set that the target signal is assumed to belong to, generated images are expected to retain empirically learned properties (Chang et al., 2017). The solution is then found by performing the following minimisation,

$$\arg \min_z \frac{1}{2} \|f(g(z)) - y\|^2. \quad (2)$$

In such a way, the solution is constrained to be within the domain of the generative model, as the recovered  $x$  is by definition generated from  $z$ , but at the same time agreement to the measurements is directly induced. Though the optimisation problem is not convex, certain theoretical bounds were derived for the cases of linear observation processes by Bora et al. (2017) and more recently phase-less linear observation processes (Hand et al., 2018).

MAP inference with generative models has been demonstrated in compressed sensing settings (Bora et al., 2017; Mardani et al., 2017) and with several common image processing tasks (Chang et al., 2017). Though it is iterative and requires the definition of a differentiable observation model, it is formally more consistent than inverse mappings; the solution is found by maximising data fidelity under empirically learned signal assumptions.

Notwithstanding their achievements, the rich classes of learning methods for CI described above focus on retrieving a single optimal solution to an inverse problem, lacking a probabilistic description of the solution space. Differently from these, we aim to infer an approximate posterior distribution from a set of measurements.

## 2.5. Conditional Variational Auto-Encoders

In our probabilistic modelling of inverse problems, we exploit conditional variational auto-encoders (CVAEs) to fit flexible posterior distributions from observations. CVAEs are latent variable generative models where the prior in the latent space  $p(z|y)$  is conditioned on an observation  $y$  (Sohn et al., 2015). As in standard VAEs, the generative model is trained with an evidence lower bound (ELBO), fitting

flexible posteriors to the training data.

CVAEs have been applied to perform a range of inference tasks, such as classification (Sohn et al., 2015), generation conditioned on classes (Nguyen et al., 2017), image-from-text inference (Yan et al., 2016) and missing value imputation (Nazabal et al., 2018). However, within CI, they have been used to infer from simple deterministic observation, such as missing pixels (Nguyen et al., 2017), while in many inverse problems the observation model is probabilistic and with potentially complicated conditional distribution.

Though we exploit CVAEs to build a flexible distribution, we aim to find a variational approximation to the posterior  $p(x|y)$  from a defined observation distribution  $p(y|x)$ , rather than building a generative model conditioned on deterministic observations. In practice, our method is different in the training of the conditional model; instead of given observations to condition the latent priors we have an observation density  $p(y|x)$  determined by our estimate of the observation process.<sup>1</sup>

## 3. A Variational Inference Model for Imaging Inverse Problems

We design a variational method to infer flexible posterior PDFs from observations in inverse problems where the prior is empirically built from image examples, hence capturing the probabilistic nature of the solution space.

### 3.1. Problem Description

We wish to recover the posterior density of signals  $p(x|y_j)$  from measurements  $y_j$ , obtained through an observation process  $y_j = f(x)$ . In CI inverse problems, such observation process is often probabilistic, as in most cases sensors readings are subject to noise and the forward model parameters are known with limited precision. Measurements can then be assumed to be drawn from a distribution  $p(y|x)$ . The form of the posterior we wish to recover is then

$$p(x|y_j) = \frac{p(y_j|x)p^*(x)}{p(y_j)}. \quad (3)$$

This distribution is intractable to directly estimate, mainly for two reasons. Firstly, in the learning setting we consider here, we do not have access to the prior distribution  $p^*(x)$ , but only to a set of signal examples which are assumed to be drawn from it  $x_n \sim p^*(x)$ . Secondly, in a general case, we can not define directly a parametric form for  $p(y_j|x)$ , but can only draw from it by stochastically generating observations as  $y_j = f_j(x)$ , where  $f_j(x)$  is drawn from the parametric family of observation processes defined

<sup>1</sup>We use the inverse problems conventional notation for signals  $x$  and observations  $y$ , which is inverted compared to the common formulation of CVAEs.

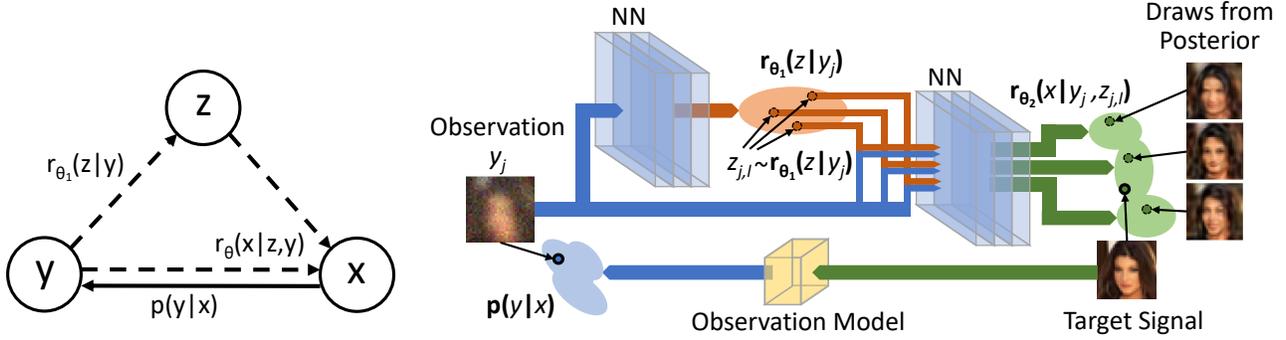


Figure 1. On the left, illustration of the conditional graphical model, where the solid line represent the true conditional dependency and the dotted lines represent variational inference. On the right, schematic diagram of the inference architecture. Measurements  $y$  are assumed to be generated from a conditional observation distribution  $p(y|x)$ . The inference model generates a latent distribution from an observation through  $r_{\theta_1}(z|y)$  and then retrieves a signal distribution  $r_{\theta_2}(x|y, z)$  conditioned on both measurements  $y$  and latent variables  $z$ .

by our estimate. For example, in a convolutional process where we are uncertain about the point spread function (PSF), we can generate an observation by first drawing a PSF  $s_j \sim p(s_j)$  and then convolving the signal  $x$  with it to obtain a draw  $y_j = f_j(x) = x \odot s_j$ . The estimate of the forward observation process and its parameters can be completely predefined by some physical model, or partially inferred from signal-observation example pairs, depending on the imaging setting and available data. Furthermore, in many cases a parametric function for  $f_j(x)$  is not available, but draws  $y_j$  can be generated through numeric simulations of the physical observation process.

### 3.2. Variational Approximation

Due to the intractability of the true posterior  $p(x|y)$ , we aim to find some parametric PDF  $r_{\theta}(x|y)$  to approximate it well over the distribution of expected measurements  $p(y)$ . To this end, we minimise the expectation of the cross entropy  $H[p(x|y), r_{\theta}(x|y)]$  under the measurements' distribution  $p(y)$ ,

$$\begin{aligned} \arg \min_{\theta} \mathbb{E}_{p(y)} H[p(x|y), r_{\theta}(x|y)] = \\ \arg \max_{\theta} \mathbb{E}_{p(y)} \int p(x|y) \log r_{\theta}(x|y) dx. \end{aligned} \quad (4)$$

Here,  $\theta$  is the set of parameters that define the distribution  $r_{\theta}(x|y)$ . The optimisation of Equation 4 is equivalent to fitting  $r_{\theta}(x|y)$  to the true posterior  $p(x|y)$  over the distribution of measurements we expect to see  $p(y)$ . We can further simplify this objective function to give

$$\begin{aligned} \mathbb{E}_{p(y)} \int p(x|y) \log r_{\theta}(x|y) dx = \\ \iint p(y) \frac{p(y|x)p^*(x)}{p(y)} \log r_{\theta}(x|y) dx dy = \\ \int p^*(x) \int p(y|x) \log r_{\theta}(x|y) dy dx. \end{aligned} \quad (5)$$

In this form, the objective function can be estimated stochastically, as the training signals  $\mathbf{x} = \{x_{1:N}\}$  are assumed to be sampled from the empirical prior  $p^*(x)$  and we can draw measurements  $y_{n,m}$  from the conditional distribution  $p(y|x_n)$  through an estimate of the observation model, as described in subsection 3.1.

### 3.3. CVAE as Approximate Posterior

As images and other natural signals often lie on complicated manifolds and the observation density  $p(y|x)$  is arbitrary, the approximate distribution  $r_{\theta}(x|y)$  needs to be rather flexible. To this end, we make use of a conditional latent variable model

$$r_{\theta}(x|y) = \int r_{\theta_1}(z|y) r_{\theta_2}(x|z, y) dz. \quad (6)$$

The latent distribution  $r_{\theta_1}(z|y)$  is an isotropic Gaussian distribution  $\mathcal{N}(z; \mu_z, \sigma_z^2)$ , where the moments  $\mu_z$  and  $\sigma_z^2$  are inferred from a measurement  $y$  by a neural network. Similarly,  $r_{\theta_2}(x|z, y)$  is an isotropic Gaussian  $\mathcal{N}(x; \mu_x, \sigma_x^2)$ , where the moments are inferred from both a measurement  $y$  and latent variable  $z$  with a second neural network. A schematic representation of the inference model is shown in Fig. 1. This approximate posterior has the same form as the marginal likelihood in CVAEs and is intractable to train directly (Sohn et al., 2015). However, we can make use of a recognition function conditioned on signals and measurements  $q_{\phi}(z|x, y)$  to derive an ELBO for  $r_{\theta}(x|y)$  that is instead efficient to estimate and optimise (Kingma & Welling, 2013; Rezende et al., 2014):

$$\begin{aligned} \log r_{\theta}(x|y) \geq \mathbb{E}_{q_{\phi}(z|x, y)} \log r_{\theta_2}(x|z, y) \\ - KL(q_{\phi}(z|x, y) || r_{\theta_1}(z|y)). \end{aligned} \quad (7)$$

The recognition function  $q_{\phi}(z|x, y)$  is also an isotropic Gaussian distribution, where the moments are inferred from

measurements  $y$  and signals  $x$  with a neural network. We can then define a lower bound for the objective function of Eq. 5 as

$$\int p^*(x) \int p(y|x) \log r_{\theta}(x|y) dy dx \geq \int p^*(x) \int p(y|x) \left[ \mathbb{E}_{q_{\phi}(z|x,y)} \log r_{\theta_2}(x|z,y) dy dx - KL(q_{\phi}(z|x,y) || r_{\theta_1}(z|y)) \right] dy dx \quad (8)$$

The above expression is equivalent to the expectation of a CVAE lower bound under the observation distribution  $p(y|x)$ , integrated over the empirical prior  $p^*(x)$ . We can estimate this bound stochastically as

$$\frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M \left[ \frac{1}{L} \sum_{l=1}^L \log r_{\theta_2}(x_n | z_{n,m,l}, y_{n,m}) - KL(q_{\phi}(z|x_n, y_{n,m}) || r_{\theta_1}(z|y_{n,m})) \right], \quad (9)$$

where  $x_n \in \mathbf{x}$  are training examples, assumed to be drawn from the signal prior  $x_n \sim p^*(x)$ ,  $y_{n,m}$  are synthetic measurements obtained from example signals  $x_n$  through a forward propagation  $y_{n,m} = f_m(x_n)$  and  $z_{n,m,l}$  are drawn from the recognition function  $q_{\phi}(z|x_n, y_{n,m})$ .

To train the approximate posterior, we maximise the empirical lower bound of Eq. 9 with respect to the parameters  $\theta_1$ ,  $\theta_2$  and  $\phi$ , over the training set  $\mathbf{x}$ . Every iteration, to draw from  $p(y|x_n)$ , we generate simulated observations  $y_{n,m}$  from a physical model of the observation process. With the trained model, we can then draw examples of solutions to the inverse problem given an observation  $y_j$  by first drawing a latent variable  $z_{j,l} \sim r_{\theta_1}(z|y_j)$  and subsequently drawing from the conditional  $x_{s,j,l} \sim r_{\theta_2}(x|z_{j,l}, y_j)$ . From these, we can compute quantities that describe the aggregate statistics of the solution space, such as pixel-wise means and error bars. A schematic representation of the model is shown in Fig. 1. The method described retains a number of desirable advantages for solving imaging inverse problems:

- From an observation, the model infers a parametric distribution of solutions, from which we can draw and obtain uncertainty measures. Such description is much richer than a single best estimate and arguably more informative for further information extraction or decision making, whether this is done by a human or an automated system.
- Inference is non-iterative, meaning that once the approximate posterior is trained, drawing from it and estimating quantities of interest is expectedly much faster than MAP methods.
- An estimate of the forward model is explicitly incorporated in the training, meaning that the approximate inversion is induced to be consistent with known properties of the physical observation process.
- While signal-observation example pairs can be used to refine or discover a forward model, if an accurate estimate is already available, only examples of the target signals are required. In many settings these are easier to obtain than observation examples, or even already available in existing data sets, avoiding the need to carry out lengthy and potentially expensive acquisitions with specific imaging instruments.

To our knowledge, the method we present here is the first to combine all the aforementioned properties. In our evaluation we make use of fully connected layers for all neural networks involved. However, the general approach we propose is compatible with other architectures, such as convolutional layers designed to process the observations in a particular setting and generative networks that model dependencies between pixels in images (Gulrajani et al., 2016).

## 4. Experiments

We carry out a number of experiments to evaluate our proposed method, both in simulation and with physical imaging systems. We first recover faces from the CelebA data set (Liu et al., 2015b) under different common simulated observation conditions, evaluating quantitatively our method. Secondly, we reconstruct images of letters deeply embedded within a scattering medium from experimental time of flight measurements. Lastly, we reconstruct images of handwritten digits and fashion items from the speckle patterns obtained when imaging through a multi-modal optical fiber.

### 4.1. Quantitative Evaluation

We make use of the CelebA dataset, down-sampled to  $32 \times 32$  images. We consider different common image observation conditions, including blurring, down-sampling and missing pixels with additive Gaussian noise, each with different parameters and parameters uncertainties. For each case, we train our variational model using 100,000 CelebA examples as training set  $\mathbf{x}$  and the corresponding observation model to generate samples  $y_{n,m} \sim p(y|x_n)$ . We then infer the approximate posterior  $r_{\theta}(x|y)$  from test observations, generated from a separate set of 20,000 images. An example showing increasingly severe blurring and the resulting inference is given in fig. 2. Additional examples and details of the experiments are given in supplementary section A. As the observed image becomes more degraded, the pixel-wise uncertainty increases and draws from the approximate posterior diversify, reflecting the broader range of likely reconstructions in the solution space. To quanti-

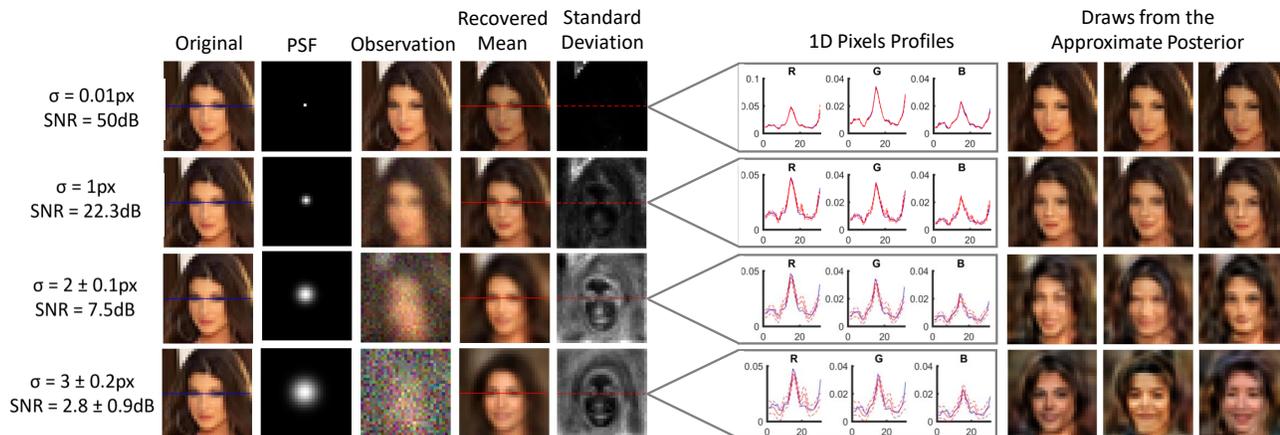


Figure 2. Inference from blurred images using our variational approach. From left to right: Original image, blurring PSF, observed image, empirical mean of the recovered approximate posterior, per-pixel empirical standard deviation (scaled and averaged over the colour channels), profiles along a 1D line across the image for each colour channel and reconstructions drawn from the approximate posterior.

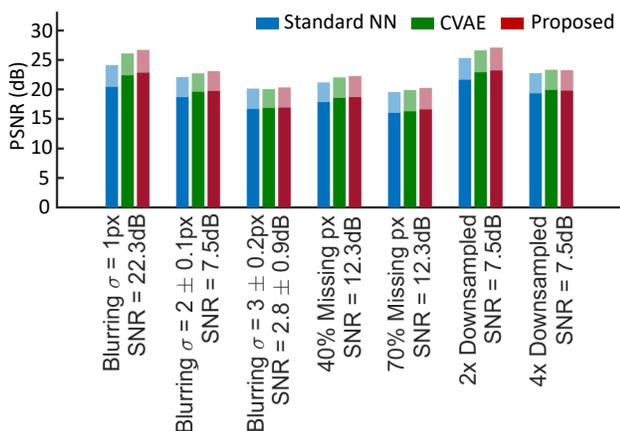


Figure 3. PSNR between original test images and mean reconstructions for standard NN, CVAE and our proposed method.

tatively evaluate the mean reconstruction performance, we compare the peak signal to noise ratios (PSNRs) obtained with our method to those of a standard neural network and a CVAE trained on simulated examples-observation pairs. As shown in Fig. 3, our proposed method proved competitive or better in every considered example. To then evaluate our method in terms of distribution accuracy, we compare the test set ELBO to that of an equivalent CVAE. As shown in table 1, in every considered case, our variational method gave an appreciably higher ELBO, proving the advantage of training through the expectation from a forward model estimate, rather than simply signal-observation pairs.

## 4.2. Imaging through Highly Scattering Media

Imaging through strongly diffusive media remains an outstanding problem in optical CI, with applications in biological and medical imaging and imaging in adverse environmental conditions. Visible or near-infrared light does

Table 1. Test set evidence lower bound for our method and a CVAE with the same neural networks.

	CVAE	Proposed
Blurring $\sigma = 1px$ SNR = 22.3 dB	14, 708	<b>15, 135</b>
Blurring $\sigma = 2 \pm 0.1px$ SNR = 7.5 dB	13, 011	<b>14, 659</b>
Blurring $\sigma = 3 \pm 0.2px$ SNR = 2.8 $\pm$ 0.9 dB	14, 226	<b>14, 547</b>
40% Missing pixels SNR = 12.3 dB	14, 517	<b>14, 752</b>
70% Missing pixels SNR = 12.3 dB	14, 393	<b>14, 626</b>
2 $\times$ Down-sampled SNR = 7.5 dB	14, 508	<b>15, 303</b>
4 $\times$ Down-sampled SNR = 7.5 dB	14, 737	<b>14, 959</b>

propagate in turbid media, such as biological tissue or fog, however, its path is strongly affected by scattering, leading to the loss of any direct image information after a short propagation length.

Following the experimental implementation presented by Boccolini et al. (2017), we employ a 130 fs near-infrared pulsed laser and a single photon sensitive time of flight (ToF) camera with a temporal resolution of 55 ps to perform transmission diffuse imaging. In these experiments, we place different cut-out shapes of alphabetic letters between two identical 2.5 cm thick slabs of diffusive material, with measured absorption and scattering coefficients of  $\mu_a = 0.09 \text{ cm}^{-1}$  and  $\mu_s = 16.5 \text{ cm}^{-1}$  respectively. We then illuminate one surface with the pulsed laser and image the opposite surface with the ToF camera. A schematic representation and a photograph of the set up are shown in fig. 5.

Signal Hidden Object	Observations			Previous MAP with $\ell_1$ +TV	Proposed Variational Method				
	Integrated Image	Time-gated Frame	Video Temporal Profile		Recovered Mean	Standard Deviation	Draws from the Approximate Posterior		
A									
X									
L									

Figure 4. Left to right: Target images embedded in the scattering medium, visualisation of the ToF camera videos used as observations, reconstruction obtained using MAP inference with  $\ell_1$ -norm and total variation regularisation and reconstruction obtained with the proposed variational method. The observed videos are shown firstly as images integrated over all time frames, then as a single temporal frame and lastly as spatio-temporal profile along one spatial dimension and time

Achieving accurate reconstructions with simple objects in this settings, is a first important step towards achieving imaging through biological tissue with near-infrared light and hence non-ionising radiation.

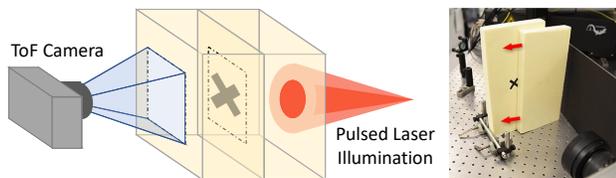


Figure 5. Left to right: schematic representation of the diffuse imaging experimental set-up and a photograph of the set up.

A pulse of light from the laser propagates through the diffusing material, reaches the hidden objects, which partially absorbs it, and then propagates further through the medium to the imaged surface. The ToF camera records a video of the light intensity as a function of time as it exits the medium. We subtract a video recorded with an empty piece of material to that obtained with the object present, thereby obtaining a video of the estimated difference in light intensity caused by the hidden object. such experimental videos are shown in Fig. 4. At this depth, more than 40 times longer than the mean free path, the diffusion effect is so severe that these basic shapes are not distinguishable directly from the videos. Furthermore, the measurements experience low signal-to-noise ratio due to the low light intensity that reaches the imaged surface and the low fill factor of the ToF camera, which is about 1%.

The reconstruction of a hidden object from a ToF video can be cast as a linear inverse problem, where the image  $x$  is observed through a linear operator  $A$  to generate an observed video  $y = Ax + \epsilon$ . Detailed descriptions of the observation model and the experiments are given in supplementary section B. This problem has been previously approached

by performing MAP inference with  $\ell_1$ -norm and TV-norm penalties (Boccolini et al., 2017). As shown in fig. 4, this technique is capable of retrieving general features of the object embedded in the scattering medium, but sometimes results in severe artefacts that make the images unrecognisable. Furthermore, to obtain the displayed results, it is necessary to carefully tune the penalty coefficients of the MAP objective function for each example, making such retrieval highly dependent on human supervision.

To approach this problem, we train our variational method with a down-sampled subset of the NIST dataset (Johnson, 2010), composed of 86, 400 hand-written numbers and letters, which we assumed to be samples from an empirical prior  $x_n \sim p^*(x)$ . Each training iteration, observations  $y_{n,m} \sim p(y|x_n)$  are generated through a simulation of the light propagation in the medium. The trained model is then used to infer an approximate posterior of solutions from the experimental ToF videos. Fig. 4 Shows the empirical mean and standard deviation of the recovered posterior, along with examples of drawn samples. Exploiting a more specific empirical prior compared to previous MAP estimate techniques, the proposed variational method retrieves more accurate reconstructions, where the different letters are clearly recognisable. Moreover, this particularly ill-posed inverse problem example highlights the importance of using a variational approach; the solution space given a diffuse video is rather variable and, unlike MAP inference and other single estimate methods, through the approximate posterior we can capture such variability by empirically estimating uncertainty and visualising different drawn samples. We further note that no training experimental acquisition was required, but only simple calibration experiments to measure the coefficients of the material. In this setting this is highly desirable as each example is manually prepared and carefully aligned before an acquisition. In addition, unlike

the analytic MAP method, the technique is non-iterative and does not require careful tuning of any parameter to obtain results such as those shown in Fig. 4.

### 4.3. Imaging Through Multi-Modal Optical Fibres

Being able to transmit images through multi-modal optical fibers (MMOFs) is a long standing goal of optical imaging and holds potential in applications ranging from endoscopy to communication. A phase coherent image can be focused into a MMOF and couple with its modes to propagate through the fibre with minimal loss of information. However, the input light field undergoes a linear complex transformation before emerging at the output as a speckle pattern (examples shown in Fig. 6). Such transformation depends on physical factors, such as fibre material, size, arrangement and even temperature. This makes the reconstruction of images from speckle patterns an extremely challenging task and traditionally requires to measure both amplitude and phase of the speckle patterns with a spectrometer (Plöschner et al., 2015).

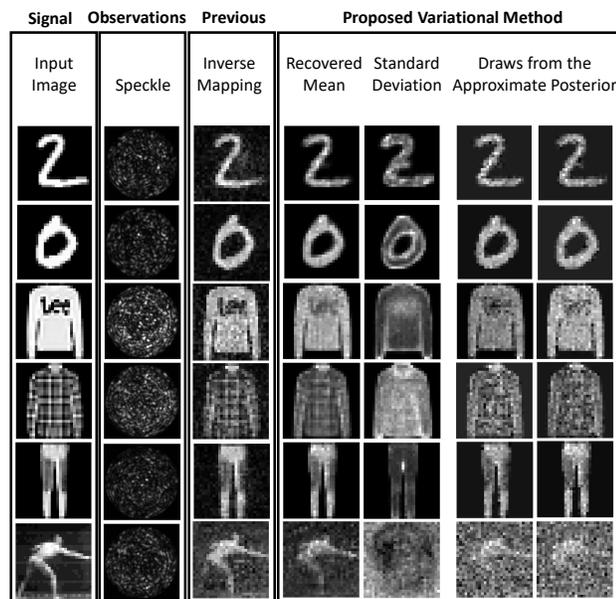


Figure 6. From left to right: test images, speckle intensities recorded at the output of the fibre, recovery using a trained inverse mapping and retrieval obtained through the proposed variational method. The bottom example differs from the training classes.

Recently, machine learning methods have been applied to learn inverse mappings from speckle intensities to images, without requiring phase measurements at the output (Moran et al., 2018; Borhani et al., 2018). We aim to perform this same task, making use of the paired images and speckles data set presented in Moran et al. (2018), but with some important differences in our approach. Instead of learning to infer images from speckle patterns directly, we use the speckles-images pairs in the training set to estimate the

coefficients of the complex transmission matrix and subsequently incorporate this learned forward observation model in the training of our approximate posterior. In such a way we are able to impose consistency with the known physical structure of the observation model while exploiting available paired examples to learn an estimate of its unknown coefficients. As in the main experiments presented in Moran et al. (2018), we train with a data set composed of images of hand-written digits, pieces of clothing and random patterns. In our probabilistic framework, this choice of training data reflects a prior which is highly “peaked” around images of numbers and clothes, but presents appreciable density everywhere, as random patterns are included in the set. In Fig. 6 We compare reconstructions obtained with our variational inference method to those recovered with the architecture presented in (Moran et al., 2018).

The average reconstruction recovered with our method is competitive with that learned through the deterministic inverse mapping. However, the variational method is capable of identifying regions of higher uncertainty, as it can be seen in the recovered standard deviations. In the clothes examples, reconstructions are less accurate in regions of fine details, but the recovered standard deviations correctly predicts such regions. In the last example, which does not resemble the training images of numbers and clothes, the reconstruction, though still recognisable, is visibly less accurate. However, this lower performance is predicted by the standard deviation returned with the variational method.

## 5. Conclusion

We presented a method to infer a variational approximation to the posterior distribution of solutions in imaging inverse problems. With such method, reconstructions are recovered non-iteratively, retain consistency with a defined physical model of the observation process and provide probabilistic information, such as pixel-wise error bars and different samples from the solution space. Simulated quantitative experiments showed superior performance to deterministic methods and standard CVAEs with equivalent neural architectures, simultaneously providing the aforementioned statistical descriptions. In real imaging experiments, the proposed method performed competitively or better in its average reconstructions compared to previous approaches and enabled new capabilities, such as predicting the accuracy of different image regions and visualising the variation in probable reconstructions through sampling. The method is also widely applicable and extendable to a variety of other systems; There are many settings, ranging from medical imaging to autonomous vehicles sensing, in which observations obey a physical model, reconstructions can greatly benefit from learning and accuracy measures are useful or even critical for decision making.

## References

- Babacan, S. D., Molina, R., and Katsaggelos, A. K. Variational Bayesian super resolution. *IEEE Transactions on Image Processing*, 20(4):984–999, 2011.
- Badon, A., Li, D., Lerosey, G., Boccara, A. C., Fink, M., and Aubry, A. Smart optical coherence tomography for ultra-deep imaging through highly scattering media. *Science advances*, 2(11):e1600370, 2016.
- Bertero, M. and Boccacci, P. *Introduction to inverse problems in imaging*. CRC press, 1998.
- Bioucas-Dias, J. M., Figueiredo, M. A., and Oliveira, J. P. Total variation-based image deconvolution: a majorization-minimization approach. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2, pp. II–II. IEEE, 2006.
- Boccolini, A., Tonolini, F., Leach, J., Henderson, R., and Faccio, D. Imaging inside highly diffusive media with a space and time-resolving single-photon sensor. In *Imaging Systems and Applications*, pp. ITu3E–2. Optical Society of America, 2017.
- Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. *arXiv preprint arXiv:1703.03208*, 2017.
- Borhani, N., Kakkava, E., Moser, C., and Psaltis, D. Learning to see through multimode fibers. *Optica*, 5:960, August 2018.
- Candes, E. J., Romberg, J. K., and Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.
- Chang, J.-H. R., Li, C.-L., Póczos, B., Kumar, B. V., and Sankaranarayanan, A. C. One network to solve them all-solving linear inverse problems using deep projection models. In *ICCV*, pp. 5889–5898, 2017.
- Daubechies, I., Defrise, M., and De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.
- Donoho, D. L. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- Durán, V., Soldevila, F., Irls, E., Clemente, P., Tajahuerce, E., Andrés, P., and Lancis, J. Compressive imaging in scattering media. *Optics express*, 23(11):14424–14433, 2015.
- Egmont-Petersen, M., de Ridder, D., and Handels, H. Image processing with neural networks – a review. *Pattern recognition*, 35(10):2279–2301, 2002.
- Figueiredo, M. A., Nowak, R. D., and Wright, S. J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007.
- Gulrajani, I., Kumar, K., Ahmed, F., Taiga, A. A., Visin, F., Vazquez, D., and Courville, A. PixelVAE: A latent variable model for natural images. *arXiv preprint arXiv:1611.05013*, 2016.
- Hand, P., Leong, O., and Voroninski, V. Phase retrieval under a generative prior. In *Advances in Neural Information Processing Systems*, pp. 9154–9164, 2018.
- Horisaki, R., Takagi, R., and Tanida, J. Learning-based imaging through scattering media. *Optics express*, 24(13):13738–13743, 2016.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. *arXiv preprint, abs/1611.07004*, 2017.
- Ji, S., Xue, Y., Carin, L., et al. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346, 2008.
- Johnson, S. G. NIST Special Database 30. 2010.
- Kingma, D. P. and Welling, M. Auto-encoding variational Bayes. *arXiv preprint arXiv*, pp. 1312.6114, 2013.
- Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., and Ashok, A. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 449–458, 2016.
- Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, volume 2, pp. 4, 2017.
- Likas, A. C. and Galatsanos, N. P. A variational approach for Bayesian blind image deconvolution. *IEEE Transactions on Signal Processing*, 52(8):2222–2233, 2004.
- Liu, F., Shen, C., and Lin, G. Deep convolutional neural fields for depth estimation from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5162–5170, 2015a.

- Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015b.
- Lucas, A., Iliadis, M., Molina, R., and Katsaggelos, A. K. Using deep neural networks for inverse problems in imaging: beyond analytical methods. *IEEE Signal Processing Magazine*, 35(1):20–36, 2018.
- Magalhães, F., Araújo, F. M., Correia, M. V., Abolbashari, M., and Farahi, F. Active illumination single-pixel camera based on compressive sensing. *Applied optics*, 50(4):405–414, 2011.
- Mardani, M., Gong, E., Cheng, J. Y., Vasanawala, S., Zaharchuk, G., Alley, M., Thakur, N., Han, S., Dally, W., Pauly, J. M., et al. Deep generative adversarial networks for compressed sensing automates mri. *arXiv preprint arXiv:1706.00051*, 2017.
- McCann, M. T., Jin, K. H., and Unser, M. A review of convolutional neural networks for inverse problems in imaging. *arXiv preprint arXiv:1710.04011*, 2017.
- Mohammad-Djafari, A. Bayesian inference tools for inverse problems. In *AIP Conference Proceedings*, volume 1553, pp. 163–170. AIP, 2013.
- Moran, O., Caramazza, P., Faccio, D., and Murray-Smith, R. Deep, complex, invertible networks for inversion of transmission effects in multimode optical fibres. In *Advances in Neural Information Processing Systems*, pp. 3284–3295, 2018.
- Nazabal, A., Olmos, P. M., Ghahramani, Z., and Valera, I. Handling incomplete heterogeneous data using vaes. *arXiv preprint arXiv:1807.03653*, 2018.
- Nguyen, A., Clune, J., Bengio, Y., Dosovitskiy, A., and Yosinski, J. Plug & play generative networks: Conditional iterative generation of images in latent space. In *CVPR*, volume 2, pp. 7, 2017.
- Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.
- Plöschner, M., Tyc, T., and Čížmár, T. Seeing through chaos in multimode fibres. *Nature Photonics*, 9(8):529, 2015.
- Rezende, D. J., Mohamed, S., and Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082*, 2014.
- Ritschl, L., Bergner, F., Fleischmann, C., and Kachelrieß, M. Improved total variation-based CT image reconstruction applied to clinical data. *Physics in Medicine & Biology*, 56(6):1545, 2011.
- Sohn, K., Lee, H., and Yan, X. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems*, pp. 3483–3491, 2015.
- Starck, J.-L., Nguyen, M. K., and Murtagh, F. Wavelets and curvelets for image deconvolution: a combined approach. *Signal processing*, 83(10):2279–2283, 2003.
- Sun, B., Edgar, M. P., Bowman, R., Vittert, L. E., Welsh, S., Bowman, A., and Padgett, M. 3D computational imaging with single-pixel detectors. *Science*, 340(6134):844–847, 2013.
- Turtaev, S., Leite, I. T., Altwegg-Boussac, T., Pakan, J. M., Rochefort, N. L., and Čížmár, T. High-fidelity multimode fibre-based endoscopy for deep-brain in vivo imaging. *arXiv preprint arXiv:1806.01654*, 2018.
- Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M. G., and Raskar, R. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3:745, 2012.
- Vogel, C. R. *Computational methods for inverse problems*, volume 23. SIAM, 2002.
- Xu, L., Ren, J. S., Liu, C., and Jia, J. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*, pp. 1790–1798, 2014.
- Yan, X., Yang, J., Sohn, K., and Lee, H. Attribute2image: Conditional image generation from visual attributes. In *European Conference on Computer Vision*, pp. 776–791. Springer, 2016.
- Yang, J. and Zhang, Y. Alternating direction algorithms for  $\ell_1$ -problems in compressive sensing. *SIAM journal on scientific computing*, 33(1):250–278, 2011.