IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 22, 2020, accepted February 15, 2020, date of publication March 4, 2020, date of current version March 18, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978435

Signal Retrieval With Measurement System Knowledge Using Variational Generative Model

ZHEYUAN ZHU[1], (Member, IEEE), YANGYANG SUN[1], JONATHON WHITE[1,2], ZENGHU CHANG[1,2], AND SHUO PANG[1]

[1]CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, FL 32816, USA
[2]Department of Physics, University of Central Florida, Orlando, FL 32816, USA

Corresponding author: Zheyuan Zhu (zyzhu@knights.ucf.edu)

This work was supported in part by the U.S. Air Force Office of Scientific Research (AFOSR) under Grant FA9550-15-1-0037 and Grant FA9550-16-1-0013, in part by the Army Research Office (ARO) under Grant W911NF-14-1-0383 and Grant W911NF-19-1-0224, in part by the Defense Advanced Research Projects Agency (DARPA) under Grant D18AC00011, and in part by the National Science Foundation under Grant 1806575.

ABSTRACT Signal retrieval from a series of indirect measurements is a common task in many imaging, metrology, and characterization platforms in science and engineering. Because most of the indirect measurement processes are well-described by physical models, signal retrieval can be solved with an iterative optimization that enforces measurement consistency and prior knowledge on the signal. These iterative algorithms are time-consuming and only accommodate a linear measurement process and convex signal constraints. Recently, neural networks have been widely adopted to supersede iterative methods by directly approximating the inverse mapping of the measurement process. However, such vanilla network with a deterministic multi-layer structure is unable to distinguish signal ambiguities in ill-posed measurement systems, and the retrieved signals often lack consistency with the measurement. In this work, we incorporate the known measurement process into a customized variational generative model to capture the distribution of all possible signals given a measurement, which can be either a linear or nonlinear process. Our signal retrieval framework resolves the ambiguity in the measurement process, and retrieves high-fidelity signals that satisfy the physical model in a variety of nonlinear, ill-posed systems, such as image retrieval from Fresnel hologram and ultrafast pulse retrieval.

INDEX TERMS Variational generative model, computational imaging, neural networks, inverse problem.

I. INTRODUCTION

Direct measurements on the signals of interest are oftentimes unavailable in many areas of science and engineering. Ingenious measurement schemes can transform the inaccessible signals to measurable quantities and facilitate the retrieval of the original signals. Many of such schemes, such as interferometry, tomography, and holography, have become standard measurement systems [1]–[4]. These measurement schemes, not necessarily following the dimension or sequence of original signals, further enable the reconstruction of abstract object dimensions [5]–[7] and engender more efficient acquisition processes [8]–[10]. Generally, the signal of interest, $\mathbf{f}$, needs to be retrieved from the measurement, $\mathbf{g}$, produced by a known measurement process $\mathbf{g} = A(\mathbf{f})$, where the forward operator $A(\cdot)$ describes the transformation model of the measurement system, and can be either linear or nonlinear. Oftentimes $A(\cdot)$ contains ambiguity. In the cases of linear model, ambiguity can be caused by an intrinsically ill-conditioned transformation, such as the compressive sensing matrix [5] or the matrix corresponding to limited-angle tomography [1]. In the case of nonlinear model, ambiguity typically arises from a nonlinear function, for example, the modulus of a complex number [3], which is not a one-to-one mapping.

Conventional signal retrieval is formulated as the task of finding an optimal reconstruction $\hat{\mathbf{f}}$ from the constrained optimization problem in Eq. (1)

$$\begin{aligned}\hat{\mathbf{f}} &= \underset{\mathbf{f}}{\operatorname{argmin}} L(\mathbf{f}) \\ &= \underset{\mathbf{f}}{\operatorname{argmin}} \left\{ \|\mathbf{g} - A(\mathbf{f})\|^2 + \lambda \boldsymbol{\phi}(\mathbf{f}) \right\}, \end{aligned} \tag{1}$$

where $\|\mathbf{g} - A(\mathbf{f})\|^2$ describes the error between the measurement from observation and the retrieved signal; the

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Asikuzzaman.

regularizer, $\boldsymbol{\phi}(\mathbf{f})$, is introduced to regularize the non-uniqueness of the ill-conditioned forward model; λ is the hyper-parameter that balances the error term and the signal regularization. The objective $L(\mathbf{f})$ can be minimized numerically with iterative algorithms [11]–[13], in which each iteration consists of two gradient descent (or proximal gradient) steps enforcing the measurement consistency and the regularization. The convergence of the iterative process requires a convex objective $L(\mathbf{f})$, which is easily satisfied for linear forward models $A(\cdot)$ and convex regularizers $\boldsymbol{\phi}(\cdot)$. However, systems that reconstruct complex-valued signal from intensity-only measurements contain either nonlinear forward operators $A(\cdot)$ or non-convex regularizers $\boldsymbol{\phi}(\cdot)$ [14], and thus suffer from stagnation or failure of the iterative algorithms [15]. Moreover, the iterative process is time-consuming, inadequate for many real-time applications.

The fast inference and the ability of learning a versatile mapping from measurement to signal contribute to the wide adoption of neural networks in recent years [16]–[20]. Most of these approaches use a neural network with parameters θ to approximate the inverse mapping $A_{inv}^{(\theta)}(\mathbf{.})$. The parameters are optimized to minimize the discrepancy between ground truth $\mathbf{f}$ and inference $A_{inv}^{(\theta)}(\mathbf{g})$ based on a series of observations $\{\mathbf{f}_i\}$. Despite its simplicity and popularity, there are two major disadvantages in such neural network inversion approach. The deterministic inversion cannot handle model ambiguity (i.e. one measurement corresponding to multiple possible signals), yielding reconstructions that resemble the average ambiguity instances in the training set [21]. In addition, the inversion network does not use the knowledge of the measurement system, and the reconstruction is usually inconsistent with the forward model [22].

The flexibility provided by the neural networks and the knowledge of the measurement systems can be combined under the Bayesian interpretation in (1). References [23]–[25] have demonstrated promising results by developing a separately trained generative model to approximate the signal prior (the regularization term), combined with the iterative algorithm for signal reconstruction. However, due to its dependency on the transpose of the image-formation process, such approach is only feasible in signal retrieval from linear forward models, and its processing time remains too long for real-time retrieval.

In this work, we incorporate the knowledge on the measurement system into a signal retrieval framework built upon conditional variational generative model. Signal retrieval process becomes a fast inference through our trained model without the use of iterations, yet it can effectively produce retrieved instances consistent with the forward models. In experiments, we demonstrate our approach in a variety of ill-posed linear and nonlinear measurement systems, including video compressive sensing (linear problem), image retrieval from Fresnel hologram (nonlinear problem), and ultrafast pulse retrieval (nonlinear problem, with phase-shift ambiguity). There had not been a single framework that is applicable to such variety of measurement processes and achieves similar or better reconstruction than the respective state-of-the-art methods. The paper is organized as follows. We first review the signal retrieval from Bayesian perspective in Section II. Then we develop our variational generative model in Section III. The experiments and the results are described in Section IV and V, respectively. Section VI concludes the paper.

II. PRELIMINARY: BAYESIAN INTERPRETATION OF SIGNAL RETRIEVAL

From Bayesian probabilistic perspective [26], the retrieved signal $\hat{\mathbf{f}}$ should be the one that maximizes the (logarithm) posterior likelihood (maximum-a-posteriori, MAP), given the measurement $\mathbf{g}$.

$$\begin{aligned}\hat{\mathbf{f}} &= \underset{\mathbf{f}}{\operatorname{argmax}} \log p(\mathbf{f}\,|\,\mathbf{g}) \\ &= \underset{\mathbf{f}}{\operatorname{argmax}} \log \frac{p(\mathbf{g}\,|\,\mathbf{f})\,p(\mathbf{f})}{p(\mathbf{g})} \\ &= \underset{\mathbf{f}}{\operatorname{argmin}} \left(-\log p(\mathbf{g}\,|\,\mathbf{f}) - \log p(\mathbf{f})\right), \end{aligned} \tag{2}$$

where $p(\mathbf{g}\,|\,\mathbf{f})$ is the likelihood of observing measurement $\mathbf{g}$ from signal $\mathbf{f}$, which is determined by both the forward process $A(\mathbf{.})$ and the noise model $p_{noise}(\mathbf{g}\,|\,A(\mathbf{f}))$ of the detection. If Gaussian noise is assumed on the detector, $p_{noise}(\mathbf{g}\,|\,A(\mathbf{f})) \sim \mathcal{N}(A(\mathbf{f}), \alpha\mathbf{I})$,

$$p(\mathbf{g}\,|\,\mathbf{f}) = C_\alpha \exp\left(-\frac{\|\mathbf{g} - A(\mathbf{f})\|^2}{\alpha}\right), \tag{3}$$

where α is the variance that reflects the Gaussian noise level of the detector, and C_α is the normalization factor. The negative logarithm of $p(\mathbf{g}\,|\,\mathbf{f})$ becomes a mean squared error (MSE) of the measurement $\|\mathbf{g} - A(\mathbf{f})\|^2$ in constrained optimization as in (1). Notice that $p(\mathbf{g}\,|\,\mathbf{f})$ can also be tailored to other detector noise models such as Poisson and Binomial etc. [12], [27]–[29]. $p(\mathbf{f})$ is the prior distribution of all plausible signals, $\mathbf{f}$. If we assume $\mathbf{f}$ follows the distribution in (4),

$$p(\mathbf{f}) = C_{\beta,\boldsymbol{\Phi}} \exp\left(-\frac{\|\boldsymbol{\Phi}(\mathbf{f})\|_p^2}{\beta}\right), \tag{4}$$

then the signal regularizer $\boldsymbol{\phi}(\mathbf{f})$ in (1) can be conceived as the negative logarithm of the prior distribution $p(\mathbf{f})$. Here the variance β determines the regularization strength, and $C_{\beta,\boldsymbol{\Phi}}$ is the normalization factor. The operator $\boldsymbol{\Phi}$ transforms $\mathbf{f}$ onto the domain $\mathbf{u} = \Phi(\mathbf{f})$, where the signal representation, $\mathbf{u}$, belongs to a simple Gaussian distribution $\mathcal{N}(\mathbf{0}, \beta\mathbf{I})$ as in (4). For compressed sensing settings based on sparsity (l^1-norm), $\boldsymbol{\Phi}$ represents the projection onto domains such as wavelet [30] or total-variation [31]. With the Gaussian distribution assumptions in (3) and (4), maximizing the posterior likelihood $p(\mathbf{f}\,|\,\mathbf{g})$ reduces to the constrained optimization problem of (1).

From the Bayesian perspective, using a generative model approach to derive a more accurate prior distribution than (4) becomes a logical follow-up [23]–[25]. The prior $p(\mathbf{f})$ was trained separately from the forward model based on a series of

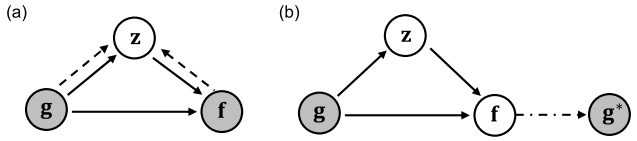


FIGURE 1. Directed graphical model (solid lines) of our proposed signal retrieval network, which contains an inference model (a) and a retrieval model (b). The signal retrieval process is parameterized by θ. Training of the parameters θ is assisted by introducing (a) variational inference process $q_\phi(\mathbf{z}|\mathbf{f},\mathbf{g})$ (dashed lines), (b) the known physical model $A(\cdot)$ of the measurement process (dot-dashed line). Variables in gray contain observable data in their respective models.

observations $\{\mathbf{f}_i\}_{i=1}^N$. Though promising retrieval results has been demonstrated, the optimization remains a lengthy iterative process. In the next section, we will describe a framework based on conditional variational generative model to directly capture the posterior distribution $p_\theta(\mathbf{f}\,|\,\mathbf{g})$, for solving various signal retrieval problems.

III. THEORY

Our signal retrieval approach implements a variational inference process that captures the posterior distribution of the signal and handles the measurement ambiguity via the introduction a latent variable $\mathbf{z}$ [20],

$$p_\theta(\mathbf{f}|\mathbf{g}) = \int p_\theta(\mathbf{f}\,|\,\mathbf{z},\mathbf{g})\, p_\theta(\mathbf{z}|\mathbf{g})\, d\mathbf{z}. \tag{5}$$

During the signal retrieval process, the latent variable $\mathbf{z}$ was sampled from the conditional prior $p_\theta(\mathbf{z}|\mathbf{g})$ given measurement $\mathbf{g}$, and the retrieved signal $\mathbf{f}$ is generated from the conditional variational distribution $p_\theta(\mathbf{f}\,|\,\mathbf{z},\mathbf{g})$. Both $p_\theta(\mathbf{z}|\mathbf{g})$ and $p_\theta(\mathbf{f}\,|\,\mathbf{z},\mathbf{g})$ distributions can be implemented with neural networks with parameter θ.

A. CONDITIONAL VARIATIONAL INFERENCE

The objective function of the variational inference model is the conditional log-likelihood $\log p_\theta(\mathbf{f}\,|\,\mathbf{g}) = \sum_{i=1}^N \log p_\theta(\mathbf{f}_i\,|\,\mathbf{g}_i)$ of the observations $\{(\mathbf{f}_i,\mathbf{g}_i), i = 1,\ldots,N\}$ with parameters θ. Due to the intractable posteriors of generative models, direct parameter estimation is generally unfeasible. However, by substituting the objective function with its variational lower bound, the parameters can be efficiently trained [32], [33]. Through the introduction of a recognition distribution $q_\phi(\mathbf{z}\,|\,\mathbf{f}_i,\mathbf{g}_i)$ (dashed lines in Figure 1(a)) as an approximation of the true posterior distribution $p_\theta(\mathbf{z}\,|\,\mathbf{f}_i,\mathbf{g}_i)$, the variational lower bound, $\mathcal{L}(\theta,\phi;\mathbf{f}_i,\mathbf{g}_i)$, can be derived as [34]

$$\begin{aligned}
&\log p_\theta(\mathbf{f}_i\,|\,\mathbf{g}_i)\\
&= E_{q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)} \log \frac{p_\theta(\mathbf{f}_i,\mathbf{z}\,|\,\mathbf{g}_i)}{q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)}\\
&\quad + KL\left(q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)\,||p_\theta(\mathbf{z}\,|\,\mathbf{f}_i,\mathbf{g}_i)\right)\\
&\geq E_{q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)} \log \frac{p_\theta(\mathbf{f}_i,\mathbf{z}\,|\,\mathbf{g}_i)}{q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)}\\
&:= \mathcal{L}(\theta,\phi;\mathbf{f}_i,\mathbf{g}_i),
\end{aligned} \tag{6}$$

where the inequality holds, because the Kullback–Leibler (KL) divergence term is always non-negative. Following the variational Bayesian approach [33], the likelihood lower bound of the inference model $\mathcal{L}$ can be expanded into

$$\begin{aligned}
\mathcal{L}(\phi,\theta;\mathbf{f}_i,\mathbf{g}_i) &= \int q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)\left(\log\frac{p_\theta(\mathbf{f}_i,\mathbf{z}|\mathbf{g}_i)}{q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)}\right)d\mathbf{z}\\
&= \int q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)\left(\log\frac{p_\theta(\mathbf{z}|\mathbf{g}_i)}{q}_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)\right.\\
&\quad \left. + \log p_\theta(\mathbf{f}_i\,|\,\mathbf{z},\mathbf{g}_i)\right)d\mathbf{z}\\
&= -KL\left(q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)\,||p_\theta(\mathbf{z}|\mathbf{g}_i)\right)\\
&\quad + E_{q_\phi(\mathbf{z}\,|\,\mathbf{f}_i,\mathbf{g}_i)}\left(\log p_\theta(\mathbf{f}_i\,|\,\mathbf{z},\mathbf{g}_i)\right).
\end{aligned} \tag{7}$$

Here we assume the conditional prior $p_\theta(\mathbf{z}\,|\,\mathbf{g})$ is a Gaussian distribution, $p_\theta(\mathbf{z}\,|\,\mathbf{g}_i) = \mathcal{N}\left(\boldsymbol{\mu}_z^{(\theta)}(\mathbf{g}_i), diag\left(\left[\boldsymbol{\sigma}_\mathbf{z}^{(\theta)}(\mathbf{g})\right]^2\right)\right)$, where the mean $\boldsymbol{\mu}_z^{(\theta)}$ and standard deviation $\boldsymbol{\sigma}_\mathbf{z}^{(\theta)}$ are implemented by neural networks. Similar assumption is applied to the recognition model, $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i) = \mathcal{N}\left(\boldsymbol{\mu}_\mathbf{z}^{(\phi)}(\mathbf{f}_i,\mathbf{g}_i), diag\left(\left[\boldsymbol{\sigma}_\mathbf{z}^{(\phi)}(\mathbf{f}_i,\mathbf{g}_i)\right]^2\right)\right)$. The KL term in $\mathcal{L}(\phi,\theta;\mathbf{f}_i,\mathbf{g}_i)$ can then be explicitly expressed as

$$\begin{aligned}
&KL\left(q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)\,||p_\theta(\mathbf{z}|\mathbf{g}_i)\right)\\
&= \sum_{j=1}^M \left(\log\frac{\sigma_{ij}^{(\phi)}}{\sigma_{ij}^{(\theta)}} + \frac{\left(\mu_{ij}^{(\theta)} - \mu_{ij}^{(\phi)}\right)^2 + \sigma_{ij}^{(\theta)^2}}{2\sigma_{ij}^{(\phi)^2}} - \frac{1}{2}\right),
\end{aligned} \tag{8}$$

where j is the index of elements in the M-dimensional vectors $\boldsymbol{\mu}_\mathbf{z}^{(\phi)}(\mathbf{f}_i,\mathbf{g}_i)$ and $\boldsymbol{\sigma}_\mathbf{z}^{(\phi)}(\mathbf{f}_i,\mathbf{g}_i)$, and their j-th elements are denoted as $\mu_{ij}^{(\phi)}$ and $\sigma_{ij}^{(\phi)}$. Similar notations are applied to $\boldsymbol{\mu}_\mathbf{z}^{(\theta)}(\mathbf{g}_i), \boldsymbol{\sigma}_\mathbf{z}^{(\theta)}(\mathbf{g}_i)$ as well. We also model $p_\theta(\mathbf{f}_i\,|\,\mathbf{z},\mathbf{g}_i)$ as a Gaussian distribution, $p_\theta(\mathbf{f}_i\,|\,\mathbf{z},\mathbf{g}_i) = \mathcal{N}(\mathbf{f}_i;\boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{z},\mathbf{g}_i),\beta\mathbf{I})$. The second term of the lower bound becomes

$$\begin{aligned}
&E_{q_\phi(\mathbf{z}\,|\,\mathbf{f}_n,\mathbf{g}_n)}\left(\log p_\theta(\mathbf{f}_n\,|\,\mathbf{z},\mathbf{g}_n)\right)\\
&\qquad \approx -\frac{1}{\beta L}\sum_{l=1}^L \left(\mathbf{f}_n - \boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{z}_l,\mathbf{g}_n)\right)^2,
\end{aligned} \tag{9}$$

where we have approximated the expectation $E_{q_\phi(\mathbf{z}\,|\,\mathbf{f}_i,\mathbf{g}_i)}$ by sampling L instances of $\mathbf{z}$ from the recognition distribution $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$ as $\{\mathbf{z}_l : l = 1,\ldots,L\}$.

B. SIGNAL RETRIEVAL WITH MEASUREMENT CONSISTENCY

During the training phase, the variational inference model draws samples $\mathbf{z}$ from the recognition distribution $q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$. The signal retrieval model, however, draws $\mathbf{z}$ from the conditional prior distribution $p_\theta(\mathbf{z}\,|\,\mathbf{g}_i)$. This inconsistency between the recognition distribution and conditional prior distribution was also recognized in Reference [32]. When using the variational lower bound as the objective function, relying only on closing the KL-divergence between $q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$ and $p_\theta(\mathbf{z}|\mathbf{g}_i)$ cannot provide effective training to the conditional prior. Here we take the measurement process

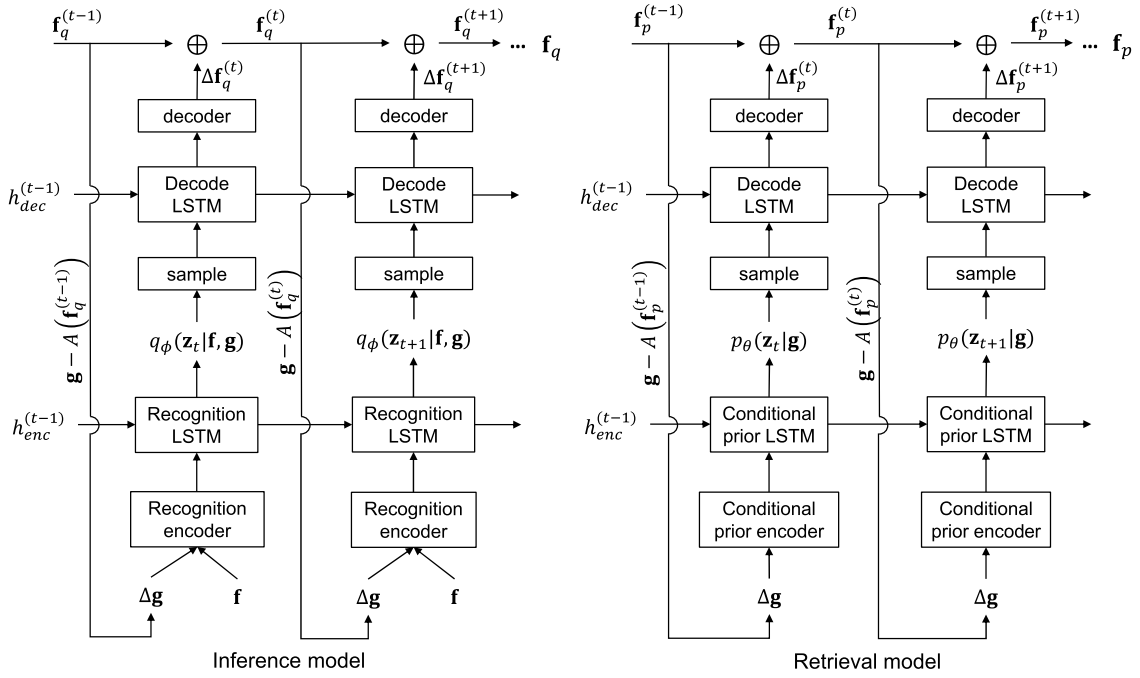


FIGURE 2. Recurrent structure of the conditional variational generative network at time stamp t. Boxes represent multi-layer structures and arrows represent data flow. In both models, the generated $\Delta\hat{\mathbf{f}}$ from all previous time stamps are aggregated to obtain f at t. The weights of the decoder are shared between inference and retrieval models.

into account and construct an alternative model to assist the training.

For the signal retrieval process, latent variable samples drawn from $p_\theta(\mathbf{z}|\mathbf{g}_i)$ capture the variance of all signals $\mathbf{f}$ that produce measurement $\mathbf{g}_i$. Naively replacing $q_\phi(\mathbf{z}|\mathbf{f}_i, \mathbf{g}_i)$ with $p_\theta(\mathbf{z}|\mathbf{g}_i)$ in the log-likelihood lower bound in (7), in an attempt to keep $\mathbf{z}$ distribution consistency, amounts to comparing all possible signals $\mathbf{f} = \boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}(\mathbf{z}_l, \mathbf{g}_i)$ given the measurement $\mathbf{g}_i$ with a single observation $\mathbf{f}_i$ in the training set. To resolve this issue, we introduce the measurement process (dot-dashed line in Figure 1(b)) to the signal retrieval model. The expected measurement, $\mathbf{g}^*$, is generated from $\boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}(\mathbf{z}, \mathbf{g}_i)$ via the forward process $\mathbf{g}^* = A\left(\boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}(\mathbf{z}, \mathbf{g}_i)\right)$. For all the $\mathbf{z}$ sampled from $p_\theta(\mathbf{z}|\mathbf{g}_i)$, we maximize the likelihood of generating the expected measurement $\mathbf{g}^*$ given point $\mathbf{g}_i$, as defined by the detection model. Applying Jensen's inequality, a lower bound of this likelihood $\mathcal{L}_r\left(\boldsymbol{\theta}; \mathbf{g}^*, \mathbf{g}_i\right)$ can be derived and used as the objective function of the retrieval process.

$$\begin{aligned}
\log p(\mathbf{g}^*|\mathbf{g}_i) &= \log \int p_\theta\left(\mathbf{g}^* \mid \mathbf{z}, \mathbf{g}_i\right) p_\theta\left(\mathbf{z}|\mathbf{g}_i\right) d\mathbf{z} \\
&\geq \int \log p_\theta\left(\mathbf{g}^* \mid \mathbf{z}, \mathbf{g}_i\right) p_\theta\left(\mathbf{z}|\mathbf{g}_i\right) d\mathbf{z} \\
&= E_{p_\theta(\mathbf{z}\mid\mathbf{g}_i)}\left(\log p_\theta\left(\mathbf{g}^* \mid \mathbf{z}, \mathbf{g}_i\right)\right) \\
&\approx -\frac{1}{\alpha L}\sum_{l=1}^{L}\left(A\left(\boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}\left(\mathbf{z}_l, \mathbf{g}_i\right)\right) - \mathbf{g}_i\right)^2 \\
&:= \mathcal{L}_r\left(\boldsymbol{\theta}; \mathbf{g}^*, \mathbf{g}_i\right), \qquad (10)
\end{aligned}$$

where we have assumed Gaussian noise model on the detector $\mathbf{g}^* \sim \mathcal{N}(\mathbf{g}_i, \alpha\mathbf{I})$. Notice that the Gaussian likelihood can be substituted with Poisson or binomial noise models in photon-limited detection [12], [28]. The expectation $E_{p_\theta(\mathbf{z}|\mathbf{g}_i)}$ in (10) is approximated by sampling L instances of from the conditional prior distribution $p_\theta\left(\mathbf{z}|\mathbf{g}_i\right)$ as $\{\mathbf{z}_l : l = 1, \ldots, L\}$. By adding in the measurement processes, we essentially construct a variational autoencoder for measurement $\mathbf{g}$, and the objective function promotes forward model consistency. We jointly train the retrieval model alongside the inference model with a hybrid objective function [32].

$$\mathcal{L}_h\left(\phi, \theta, \mathbf{f}_i, \mathbf{g}_i\right) = \gamma\mathcal{L}\left(\phi, \theta; \mathbf{f}_i, \mathbf{g}_i\right) + (1-\gamma)\,\mathcal{L}_r\left(\theta; \mathbf{g}^*, \mathbf{g}_i\right), \qquad (11)$$

where the hyperparameter γ balances the weight between the two models.

IV. EXPERIMENTS

The conditional variational generative model consisted of both inference and retrieval models, as indicated in (11). The implementations of the two models are detailed in Figure 2. The encoder of the inference model takes in both $\mathbf{f}$ and $\mathbf{g}$ as inputs. The encoder of the retrieval model only accepts one input, $\mathbf{g}$. The mapping from latent domain to the signal domain $\mathbf{F}$ was performed by a decoder whose weights are shared in both models. Inspired by the conventional iterative algorithms, our network adopted a recurrent construction [35] with LSTM encoders and decoders, whose states at recurrence t are denoted as $h_{enc}^{(t)}$ and $h_{dec}^{(t)}$, respectively. Outputs from the both models, $\mathbf{f}_q$ and $\mathbf{f}_p$, were initialized as 0. Each recurrence generated an increment $\Delta\mathbf{f}^{(t)}$ to $\mathbf{f}$ from the discrepancy $\mathbf{g} - A\left(\mathbf{f}^{(t-1)}\right)$ between observed measurement $\mathbf{g}$ and

the previous estimate $A\left(\mathbf{f}^{(t-1)}\right)$. For the first recurrence, this discrepancy was set to $\mathbf{g}$. During the signal retrieval process, one measurement $\mathbf{g}_i$ from the test dataset and one sample $\mathbf{z}_l \sim p_\theta(\mathbf{z}|\mathbf{g}_i)$ are fed into the generative network $p_\theta(\mathbf{f}|\mathbf{g}_i, \mathbf{z}_l)$ to obtain one reconstruction instance $\hat{\mathbf{f}}_i$. As a comparison, we also trained a single-pass deterministic neural network based on the structure of the retrieval model. The sampling process and physical model were removed in the deterministic network. The loss function of the deterministic network consisted only of the MSE on $\mathbf{f}$. All the networks and physical model of the measurement process were implemented in TensorFlow 1.9.0 and Python 3.6 environment.

The reconstruction performance is evaluated by peak signal-to-noise ratio (PSNR), which is defined as PSNR$=10\times\log_{10}(\max(\mathbf{f}_i)/\text{MSE})$, where MSE$=\frac{1}{\dim(\mathbf{f}_i)}\left\|\hat{\mathbf{f}}_i-\mathbf{f}_i\right\|^2$ is the mean square error between the ground truth $\mathbf{f}_i$ and the reconstructed instance $\hat{\mathbf{f}}_i$ across all dimensions of $\mathbf{f}$. The fidelity, defined as the PSNR between the measurements generated from reconstruction $\mathbf{g}^* = A(\hat{\mathbf{f}}_i)$ and ground truth $\mathbf{g}_i$, quantifies how well the reconstructions match the physical model of the measurement process. We compared the reconstruction performance between our model and deterministic networks with three signal retrieval examples, detailed in the following subsections.

A. CODED APERTURE VIDEO COMPRESSIVE SENSING

Video compressive sensing encodes fast-moving scenes with alternating masks on the conjugate image plane so that they can be captured by a slow camera [5]. Each low-frame-rate measurement recorded on the camera, $I(x, y)$, is the high-speed scene $f(x, y, t)$ encoded by a series of rapidly-changing mask $M(x, y, t_i)$

$$I(x,y)=\int_{t=0}^{K\tau} f(x,y,t)\sum_{i=1}^{K} M(x,y,t_i)rect(\frac{t-t_i-\tau/2}{\tau})dt, \tag{12}$$

where $1/\tau$ is the frequency of the changing mask. The frame rate of the camera, $1/(K\tau)$, is K times slower than the mask frequency.

In this example, we demonstrate the compression of $K = 4$ color frames into 1 measurement with random binary masks. The number of pixels in both the high-speed scene and measurement were $N \times N$ ($N = 64$). The spatial-encoding binary masks apply to all color channels, and are represented by a Kronecker product $\mathbf{M}_i = \mathbf{m}_i \otimes \mathbf{1}^{1\times1\times3}, i = 1, 2, 3, 4$, where $\mathbf{m}_i \in \{0, 1\}^{N\times N}$ denotes the transmittance of the mask, and $\mathbf{1}^{1\times1\times3}$ is a unit tensor along the dimension of color channels. Let $\{\mathbf{f}_i \in \mathbb{R}^{N\times N\times3}, i = 1, 2, 3, 4\}$ denote the 4 color frames from the fast-moving scene within one measurement frame. The measurement $\mathbf{g} \in \mathbb{R}^{N\times N\times3}$ is given by $\mathbf{g} = \sum_{i=1}^{4}\mathbf{M}_i \odot \mathbf{f}_i$, where $\odot$ denotes the element-wise product between tensors. This measurement process can be described by a linear forward model $\mathbf{g} = \mathbf{A}\mathbf{f}$, in which $\mathbf{f}$ and $\mathbf{g}$ are vectorized into $\mathbb{R}^{12N^2}$ and $\mathbb{R}^{3N^2}$ respectively; $\mathbf{A}$ is concatenated from four diagonal matrices $[diag(\mathbf{M}_1), diag(\mathbf{M}_2), diag(\mathbf{M}_3), diag(\mathbf{M}_4)]$, where $\mathbf{M}_i$ is vectorized into $\mathbb{R}^{3N^2}$.

The network for video compressive sensing employed convolutional layers in encoders, LSTM cells and decoders. The number of recurrences was 3. The network was trained on random four-image combinations from the ImageNet database, and tested on 100 traffic video segments in DynTex library.

B. IMAGE RETRIEVAL FROM FRESNEL HOLOGRAM

We constructed a Fresnel in-line hologram forward model based on the setup in [36]. Coherent, parallel beam illumination ($\lambda = 635$nm) was assumed in the forward model and the propagation distance z between the object and the detector plane was set to 400mm. The intensity on the camera is the interference between the propagated field and the reference beam

$$I(x,y)=\left|A_{ref}+\tilde{E}_{prop}(x,y)\right|^2, \tag{13}$$

where A_{ref} represents the parallel, on-axis reference field. The complex field, $\tilde{E}_{prop}$, is given by the Fresnel propagation of incident field $\tilde{E}_o$,

$$\tilde{E}_d(x,y)=\int \tilde{E}_o(x_0,y_0)\exp\left[\frac{\mathbb{i}\pi}{z\lambda}\left((x-x_0)^2+(y-y_0)^2\right)\right]dx_0dy_0, \tag{14}$$

where (x_0, y_0) is the spatial coordinates of the incident field, and z is the propagation distance.

This example considers the retrieval of a real object from its in-inline Fresnel hologram. The fields on the object $\tilde{E}_o \in \mathbb{C}^{64\times64}$ and camera plane $\tilde{E}_d \in \mathbb{C}^{64\times64}$ were both discretized into 64×64 pixels, with a pixel size of 50μm. The input of the forward model, $\mathbf{f} \in \mathbf{R}^{64\times64}$, was a zero-padded MNIST digit representing the real part of $\tilde{E}_o$. The imaginary part of $\tilde{E}_o$ was set to zero. Let $\mathbf{x}_0$, $\mathbf{y}_0$ denote the coordinates of pixels on the object plane, the complex field on the camera plane, $\tilde{E}_d$, can be formulated as a two-dimensional, discrete convolution between input field $\tilde{E}_o$ and a quadratic phase kernel $\tilde{F} = \exp\left[\frac{\mathbb{i}\pi}{z\lambda}\left(\mathbf{x}_o^2+\mathbf{y}_o^2\right)\right]$. The measured intensity, $\mathbf{g}$, is the squared modulus of the complex field $\tilde{E}_d$. As a result, we adopted convolutional structures in the encoders, LSTM cells and decoder. The number of recurrences was 2. The network was trained on Fresnel holograms simulated from 10000 MNIST training digits for 40 epochs, and tested on 1000 pairs of holograms and digits from the MNIST test dataset.

C. ULTRAFAST PULSE RETRIEVAL

In this example, we retrieve the amplitude and phase of ultrafast laser pulses from the streaking trace. The forward process of streaking was established based on the theory in Reference [19]. The streaking trace is a series of photoelectron spectra arising from the interaction between an attosecond

extreme ultraviolet (XUV) pulse $\tilde{\vec{E}}_{XUV}(t)$ and a femtosecond infrared (IR) dressing field $\tilde{\vec{E}}_{IR}(t)$ under different time delays τ. In atomic units, the streaking intensity $I(K,\tau)$ is

$$I(K,\tau) = \left| \int \tilde{\vec{E}}_{XUV}(t-\tau)\cdot\vec{d}\exp \mathbb{i}\phi_G(K,t)\exp(-\mathbb{i}(K+I_p)t)dt \right|^2, \quad (15)$$

where $\mathbb{i}$ is the imaginary unit, K is the kinetic energy of the photoelectron; I_p is the ionization potential; $\vec{d}$ is the dipole transition matrix element from the ground state to the continuum state [4], and is assumed to be constant [37]; ϕ_G is a phase gate on the photoelectron wave $\tilde{\vec{E}}_{XUV}(t-\tau)\cdot\vec{d}$, and is determined by the IR dressing field via [37]

$$\phi_G(K,t) = -\int_t^{\infty}\left[\vec{v}\cdot\vec{A}(t') + \frac{\vec{A}^2(t')}{2}\right]dt', \quad (16)$$

where $\vec{v}$ is the momentum of the electron and is related to the kinetic energy via $K = \vec{v}^2/2$; $\vec{A}(t) = -\partial\tilde{\vec{E}}_{IR}/\partial t$ is the vector potential of the IR dressing field. In streaking experiments, the X-ray and IR fields are linearly polarized along the same directions, such that $\tilde{\vec{E}}_{XUV}$ and $\tilde{\vec{E}}_{IR}$ could both be reduced to scalar field $\tilde{E}_{XUV}$ and $\tilde{E}_{IR}$.

The XUV and IR pulses were created by imposing the experimental XUV and IR spectra on their corresponding spectral phases $\tilde{E}(\omega) = \sqrt{S(\omega)}\exp \mathbb{i}\phi(\omega)$, and Fourier-transformed into the time domain for streak calculation. The spectral phase term was expressed as a 5th order polynomial function $\phi(\omega) = \sum_{i=0}^{5} k_i\omega^i$. The number of sampling for XUV and IR spectra $\tilde{\mathbf{E}}_{XUV}$ and $\tilde{\mathbf{E}}_{IR}$ was 200 and 20, respectively. The input of this forward model, $\mathbf{f} \in \mathbb{R}^{440}$, was a concatenated vector representing the real and imaginary part of the XUV and IR spectra, $[\mathrm{Re}\left(\tilde{\mathbf{E}}_{XUV}\right), \mathrm{Im}\left(\tilde{\mathbf{E}}_{XUV}\right), \mathrm{Re}\left(\tilde{\mathbf{E}}_{IR}\right), \mathrm{Im}(\tilde{\mathbf{E}}_{IR})]$. The output of the forward model, $\mathbf{g} \in \mathbb{R}^{256\times 35}$, was the discretized streaking trace I in terms of 256 energies K ranging from 50 to 305 eV, and 35 time delays τ from -8fs to 8fs. Since the carrier envelop phase (CEP) term k_0 of the XUV pulse does not affect the streaking intensity, XUV pulses with the same phase coefficients except k_0 would yield identical streak traces $\mathbf{g}$, creating ambiguities in the training dataset.

The network for ultrafast pulse retrieval consisted of convolutional encoders, fully-connected LSTM cells and decoder. The number of recurrences was 2. The network was trained on streak traces simulated from 10000 XUV and IR pulses with random phase coefficients k_0 to k_5 for 100 epochs. The test dataset contained another 100 streak traces. For each streak trace, we sampled 10 reconstruction instances from the approximated posterior distribution $p_\theta(\mathbf{f}|\mathbf{g})$.

V. RESULTS AND DISCUSSIONS

A. CODED APERTURE VIDEO COMPRESSIVE SENSING

We first demonstrated that our network is equivalent to a learning-based signal prior for a linear imaging process. The network was trained on 4X coded aperture compression forward model and tested on compressed video frames. Figure 3 (a) shows the ground truth of the 4 frames and the compressive measurement. The 4 reconstructed frames are shown in Figure 3 (b), along with the compressive measurement from the reconstructed frames. As a comparison, we also performed iterative maximum-a-posteriori reconstructions with TV prior and deep pixel-level prior (DPP) [23], shown in Figure 3 (c-d), respectively. The optimized strength of TV prior was $10^{2.0}$. The optimal step parameter of the DPP network, analogous to the regularization strength, was 0.5. The PSNR and fidelity of TV, DPP and our model are listed in Table 1. We speculate the lower fidelity is attributed to having only 3 recurrences in our model, which is currently limited by the computation power of GPU. However, it is worth noting that the reconstruction time (0.13s) of our trained model was orders of magnitude shorter than DPP, which required hundreds of iterations and took 197.3s. Both our model and DPP outperform TV thanks to their more realistic prior distributions.

TABLE 1. PSNR and fidelity of the reconstructions using TV, DPP and our method.

	TV	DPP	Ours
PSNR	22.07	22.44	22.14
Fidelity	37.05	37.12	24.43

B. IMAGE RETRIEVAL FROM IN-LINE FRESNEL HOLOGRAM

In this experiment, we demonstrate the performance of our model in retrieving the image from Fresnel hologram, which is a nonlinear image formation process. The Fresnel holograms (Figure 4 (a2)) simulated from the MNIST test images (Figure 4 (a1)) were fed into the deterministic, physics-informed and our model trained on the holograms simulated from MNIST training dataset. The reconstructions were then forward propagated to the detector plane to evaluate the fidelity. We also performed reconstructions from a deterministic network, and a physics-informed network [36] that adds a Fresnel back-propagation operation before the deterministic network. Table 2 lists the PSNR and fidelity of all the test images retrieved from deterministic network, physics-informed network and our model with comparable structures. The fidelity of our model is better than both deterministic and physics-informed neural networks. Though physics-informed network embeds the Fresnel back-propagation as its first layer, the back-propagated image still suffers from the twin image artifact, which needs to be corrected by the subsequent deterministic neural network. In our model, we apply the Fresnel forward propagation to the intermediate reconstruction and feed the error of the measurement back into the encoder, thus achieving a higher fidelity via direct enforcement of the forward model on the reconstructed image.

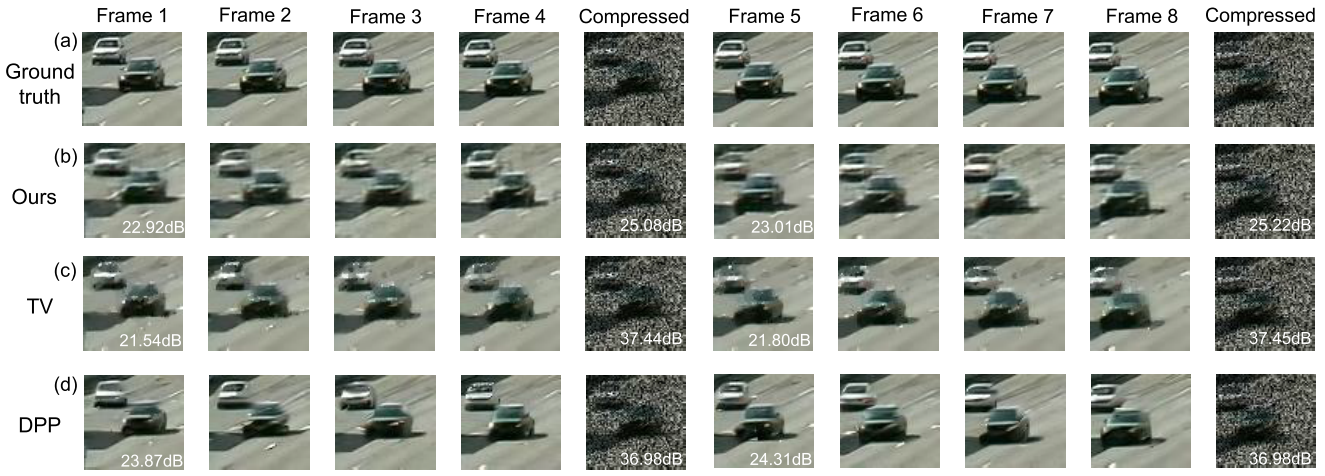


FIGURE 3. Reconstructed frames from 4X video compression model: (a) Ground truth of the frames and simulated compressed measurement. (b-d) Reconstructed frames and measurements from our model, TV and Deep Pixel-level Prior (DPP), respectively. The number on the first frame indicates the PSNR of all 4 retrieved frames. The number on the compressed measurement indicates the fidelity of the measurement calculated from 4 retrieved frames.

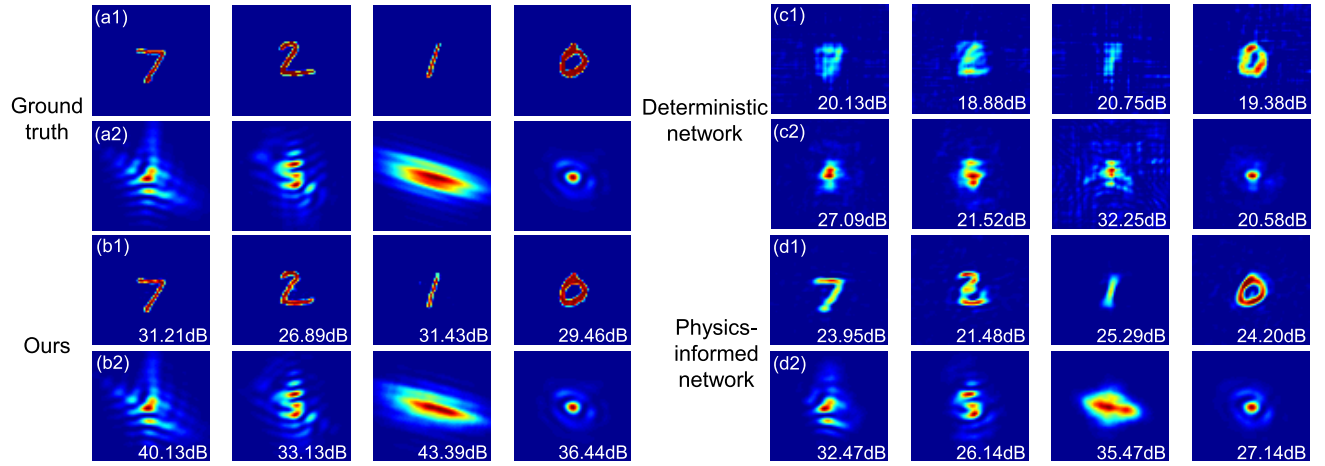


FIGURE 4. Reconstructed images from Fresnel hologram: (a) Ground truth of the image (a1) and simulated Fresnel hologram intensity (a2). (b-d) Reconstructed images and holograms from our model, deterministic network, and physics-informed network.

TABLE 2. PSNR and fidelity of the reconstructed image from holograms with deterministic network, physics-informed network and our method.

	Deterministic network	Physics-informed network	Ours
PSNR	19.35	22.83	27.21
Fidelity	23.46	28.86	35.30

C. ULTRAFAST PULSE RETRIEVAL

Finally, we demonstrate the capability of our model to resolve ambiguities of a nonlinear forward model in the ultrafast pulse retrieval example. Figure 5 displays the real and imaginary part of the XUV spectrums retrieved from the input streaking trace, along with reconstructed traces simulated from the forward process $A(\cdot)$. An XUV pulse in test dataset (Figure 5 (a1)) produced the streaking trace in Figure 5(a2), which was fed into the trained signal retrieval process. Three instances of the retrieved XUV spectrums from Figure 5 (a2) are shown in Figure 5 (b1-d1), with PSNR of 5.68, 10.65 and 15.85dB, respectively, compared with the ground truth in Figure 5 (a). Figure 5 (b3-d3) show the traces reconstructed from retrieved pulses (Figure 5 (b1-d1)), and the fidelity with respect to the ground truth in Figure 5(a2).

The high measurement fidelities suggest that instances in Figure 5 (b1-d1) belong to the phase-shift ambiguities of the same streaking trace. To verify this, we shifted the CEP term k_0 of the retrieved XUV spectrums by their average phase differences between Figure 5 (a1) in the energy range 100∼300eV. The resulting spectrums (Figure 5 (b2-d2)) all match the ground truth with good consistency. The amounts of CEP shift were 1.65, 0.84 and −0.39 radians, respectively, with PSNR of 26.99, 26.77 and 21.67 after the phase shift. In contrast, the deterministic network generates identical reconstructions similar to the average of the ambiguity instances. The XUV spectrum in Figure 5 (e1) cannot be phase-shifted to match the ground truth, and exhibits low fidelity (Figure 5 (e3)) compared with

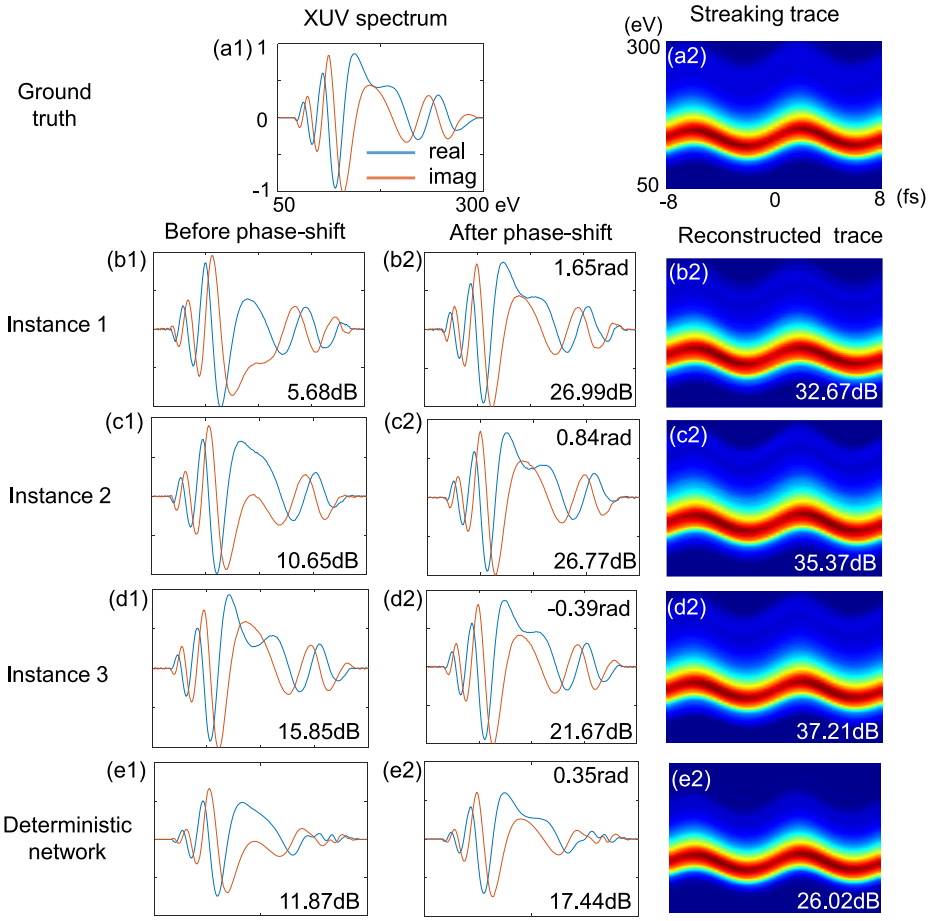


FIGURE 5. Reconstructions from the ultrafast pulse retrieval experiment: (a) Ground truth of the real and imaginary part of the XUV spectrum and its simulated streak trace. The IR spectrum is not shown in the figure. (b-d) Three instances of retrieved XUV spectrum (b1-d1), their phase-shifted variant (b2-d2), and the streak trace (b3-d3) calculated from each instance. (e) Retrieved XUV spectrum, its phase-shifted variant and streak trace from the deterministic network.

TABLE 3. PSNR and fidelity of the reconstructed pulse with deterministic network and our method.

	Deterministic network	Ours
PSNR	11.08	13.87
Fidelity	23.10	31.49

the actual measurement. Table 3 summarizes the average PSNR and fidelity of the reconstructions from the 100 test streak traces, each generating 10 instances of XUV spectrums. The high fidelity of our method indicates that it can generate different reconstruction instances satisfying the measurement forward model, a capability not possessed by deterministic network. To reach similar reconstruction fidelity from a deterministic network, manual removal of the ambiguity instances from the training data is required.

VI. CONCLUSION

We have proposed a model-based conditional generative network for solving a wide variety of linear and non-linear signal retrieval problems, including coded aperture video compressive sensing, image retrieval from Fresnel hologram and ultrafast pulse retrieval. The proposed framework exploits the known forward process of the measurement systems to train the conditional variational generative model. Compared with deterministic neural network that approximates the inversion of the forward process, our variational generative network resolves ambiguities in the training dataset, and demonstrates high-fidelity reconstructions that are consistent with the measurement process for both linear and non-linear forward models. We envision our framework as a general signal retrieval pipeline for a variety of measurement processes in which the indirect measurement obeys a physical model.

APPENDIX

The implementation of the proposed model is available at https://github.com/zyzhucreol/signal_retrieval_cvae.

REFERENCES

[1] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Philadelphia, PA, USA: SIAM, 2001.

[2] J. W. Goodman and R. W. Lawrence, "Digital image formation from electronically detected holograms," *Appl. Phys. Lett.*, vol. 11, no. 3, pp. 77–79, Aug. 1967.

[3] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.

[4] J. Itatani, F. Quéré, G. L. Yudin, M. Y. Ivanov, F. Krausz, and P. B. Corkum, "Attosecond streak camera," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 173903.

[5] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," *Opt. Express*, vol. 21, no. 9, p. 10526, Apr. 2013.

[6] G. Harding, J. Kosanetzky, and U. Neitzel, "X-ray diffraction computed tomography," *Med. Phys.*, vol. 14, pp. 515–525, Jul. 1987.

[7] W. L. Wolfe, *Introduction to Imaging Spectrometers*, vol. 25. Bellingham, WA, USA: SPIE, 1997.

[8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[9] Z. Zhu, R. A. Ellis, and S. Pang, "Coded cone-beam X-ray diffraction tomography with a low-brilliance tabletop source," *Optica*, vol. 5, no. 6, pp. 733–738, Jun. 2018.

[10] T.-H. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Spectral-temporal compressive imaging," *Opt. Lett.*, vol. 40, no. 17, pp. 4054–4057, Aug. 2015.

[11] J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new TwIST: Two-step iterative Shrinkage/Thresholding algorithms for image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2992–3004, Dec. 2007.

[12] Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—Theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[14] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 19, no. 7, pp. 1334–1345, Jul. 2002.

[15] J. R. Fienup and C. C. Wackerman, "Phase-retrieval stagnation problems and solutions," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 3, no. 11, p. 1897, Nov. 1986.

[16] R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Opt. Express*, vol. 24, no. 13, p. 13738, Jun. 2016.

[17] A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica*, vol. 4, no. 9, p. 1117, Sep. 2017.

[18] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[19] J. White and Z. Chang, "Attosecond streaking phase retrieval with neural network," *Opt. Express*, vol. 27, no. 4, p. 4799, Feb. 2019.

[20] J. Zhao, Y. Sun, Z. Zhu, J. E. Antonio-Lopez, R. A. Correa, S. Pang, and A. Schülzgen, "Deep learning imaging through fully-flexible glass-air disordered fiber," *ACS Photon.*, vol. 5, no. 10, pp. 3930–3935, Sep. 2018.

[21] C. Doersch, "Tutorial on variational autoencoders," 2016, *arXiv:1606.05908*. [Online]. Available: http://arxiv.org/abs/1606.05908

[22] F. Tonolini, J. Radford, A. Turpin, D. Faccio, and R. Murray-Smith, "Variational inference for computational imaging inverse problems," 2019, *arXiv:1904.06264*. [Online]. Available: http://arxiv.org/abs/1904.06264

[23] A. Dave, A. K. Vadathya, R. Subramanyam, R. Baburajan, and K. Mitra, "Solving inverse computational imaging problems using deep pixel-level prior," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 37–51, Mar. 2019.

[24] J. H. R. Chang, C.-L. Li, B. Poczos, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "One network to solve them all—Solving linear inverse problems using deep projection models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5889–5898.

[25] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, "Unrolled optimization with deep priors," 2017, *arXiv:1705.08041*. [Online]. Available: https://arxiv.org/abs/1705.08041

[26] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, vol. 3. Philadelphia, PA, USA: SIAM, 2006.

[27] A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," *Science*, vol. 343, pp. 58–61, Jan. 2014.

[28] Z. Zhu and S. Pang, "Few-photon computed X-ray imaging," *Appl. Phys. Lett.*, vol. 113, no. 23, Dec. 2018, Art. no. 231109.

[29] Z. Zhu, H.-H. Huang, and S. Pang, "Photon allocation strategy in region-of-interest tomographic imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 125–137, Jun. 2019.

[30] B. Zhang, J. M. Fadili, and J. L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1093–1108, Jul. 2008.

[31] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[32] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 3483–3491.

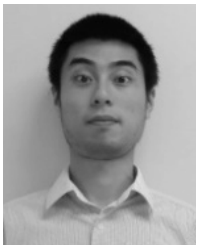
[33] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*. [Online]. Available: https://arxiv.org/abs/1312.6114

[34] C. M. Bishop, *Pattern Recognition and Machine Learning*. Cham, Switzerland: Springer, 2006.

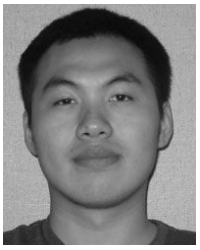
[35] K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "DRAW: A recurrent neural network for image generation," May 2015, *arXiv:1502.04623*. [Online]. Available: https://arxiv.org/abs/1502.04623

[36] A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Phys. Rev. Lett.*, vol. 121, no. 24, Dec. 2018, Art. no. 243902.

[37] Y. Mairesse and F. Quéré, "Frequency-resolved optical gating for complete reconstruction of attosecond bursts," *Phys. Rev. A, Gen. Phys.*, vol. 71, no. 1, pp. 1–4, Jan. 2005.



ZHEYUAN ZHU (Member, IEEE) received the B.S. degree in physics from Nanjing University, Nanjing, China. He is currently pursuing the Ph.D. degree in optics and photonics at CREOL, The College of Optics and Photonics, University of Central Florida (UCF). His research focuses on designing, modeling, and prototyping novel computational imaging platforms using minimal system resources in both visible and X-ray regimes, with a specialty in X-ray transmission/diffraction tomography. He was a recipient of CREOL Student of the Year Finalist Award, UCF Graduate Research Support Award, as well as various other travel grants. He also served as the Vice President of the IEEE Photonics Society student chapter at CREOL and the President of the CREOL Association of Optics Students.

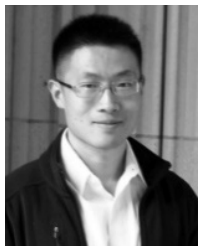


YANGYANG SUN received the B.S. degree in information engineering from Nanjing University, China, and the Ph.D. degree in optics from CREOL, The College of Optics and Photonics, University of Central Florida.

His research focuses on computational imaging and deep learning for prototyping novel imaging systems and sensing strategies. He was a recipient of the ORC Graduate Fellowship. He also served as the President for the SPIE Student Chapter at CREOL.

JONATHON WHITE received the B.Sc. degree in engineering physics and mechanical engineering from Kettering University, Michigan. He is currently pursuing the M.Sc. degree with the Abbe School of Photonics (Jena). He has conducted research at the Institute for the Frontier of Attosecond Science and Technology (iFAST, University of Central Florida) and the Institute of Applied Physics (Friedrich Schiller University Jena) in the field of machine learning applied to attosecond optics and coherent diffractive imaging.

ZENGHU CHANG received the bachelor's degree from Xi'an Jiaotong University, China, in 1982, the master's and Ph.D. degrees from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, in 1985 and 1988, respectively. He is currently a University Trustee Chair, Pegasus, and a Distinguished Professor at the University of Central Florida (UCF), where he directs the Institute for the Frontier of Attosecond Science and Technology. In 2010, he started the Joined Faculty Position in CREOL and the Department of Physics, UCF. He is the author of the book *Fundamentals of Attosecond Optics*. His notable contributions include the development of attosecond X-ray sources in the Water Window. He is a Fellow of the American Physical Society and Optical Society of America.



SHUO PANG received the B.S. degree in optical engineering from Tsinghua University, Beijing, China, the M.S. degree in biomedical engineering from the Texas A&M University, College Station, Texas, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, California.

He was a Postdoctoral Associate with the Department of Electrical and Computer Engineering, Duke University, from 2013 to 2014. He is currently an Assistant Professor at CREOL, The College of Optics and Photonics, University of Central Florida. His current research focuses on developing computational imaging systems, image processing in both visible and X-ray regimes, and machine-learning approach in optics.

He was a recipient of the Ralph E. Powe Junior Faculty Award, in 2016 and the SPIE Defense and Commercial Sensing (DCS) Rising Researcher Award, in 2017. He was the Chair of the Optical Microscopy Group of the Optical Society of America. He serves on the organizing committee of Anomaly Detection and Imaging with X-ray Conference of SPIE DCS and a Co-Editor for the CREOL Institutional Focus Issue in Applied Optics, in 2019.