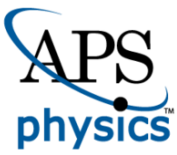

Low Photon Count Phase Retrieval Using Deep Learning

Alexandre Goy, Kwabena Arthur, Shuai Li, and George Barbastathis

Phys. Rev. Lett. **121**, 243902 — Published 12 December 2018

DOI: 10.1103/PhysRevLett.121.243902

Low Photon Count Phase Retrieval Using Deep Learning

Alexandre Goy,[1,*] Kwabena Arthur,[1] Shuai Li,[1] and George Barbastathis[1,†]

[1] *Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA02139, USA.*

(Dated: November 13, 2018)

Imaging systems' performance at low light intensity is affected by shot noise, which becomes increasingly strong as the power of the light source decreases. In this paper we experimentally demonstrate the use of deep neural networks (DNNs) to recover objects illuminated with weak light and demonstrate better performance than with the classical Gerchberg-Saxton phase retrieval algorithm for equivalent signal over noise ratio. The prior contained in the training image set can be leveraged by the DNN to detect features with a signal over noise ratio close to one. We apply this principle to a phase retrieval problem and show successful recovery of the object's most salient features with as little as one photon per detector pixel on average in the illumination beam. We also show that the phase reconstruction is significantly improved by training the neural network with an initial estimate of the object, as opposed to training it with the raw intensity measurement.

Many imaging systems only yield partial or distorted information about the object being imaged. Typical causes include loss of spatial frequencies, lack of phase information, unknown scatterers in the optical train, aberrations, and noise in the illumination or detection. In these situations, the mathematical operator describing the imaging system becomes ill-posed and usually requires regularization. A regularizer is an operator designed to favor solutions that match our prior knowledge about the object, if any. The choice of the regularizer itself is often arbitrary and based on practical experience. Recently, Deep Neural Networks (DNNs) have attracted much attention in the field of computational imaging, for they provide a way to regularize a problem adaptively. As of today, DNNs have been proven efficient solvers in many imaging applications such as deblurring [1], undersampled imaging [2], ghost imaging [3], phase retrieval [4–9], adaptive illumination microscopy [10], adaptive optics [11], and optical tomography [12, 13]. For consumer cameras operating with broadband, spatially incoherent light of flux as low as $\sim$0.1 lx at the camera, a DNN can recover images with significant detail [14]; also for phase retrieval with coherent illumination, numerical results show that DNNs outperform classical methods on noisy data [7].

In this paper, we demonstrate experimentally for the first time, to our knowledge, that DNNs can solve a coherent phase retrieval problem affected by strong shot noise at various levels. In situations where the light source is weak, the detection signal to noise ratio (SNR) is ultimately limited by the quantized nature of light. Because of its fundamental nature, shot noise cannot be avoided and regularization schemes must be devised to handle it. As the noise becomes more significant, reconstruction algorithms' performance in general deteriorates; this is the regime where we expect the biggest payoff from the DNN, assuming that it has been successfully trained to recover the object features that best explain the observed signal distribution. Best results are obtained for objects within restricted classes, *i.e.* sharing similar constrained features, or equivalently having a sparse description in some domain of appropriately chosen basis functions. To illustrate this, we used two sets of databases to train DNNs: a relatively restricted class of Integrated Circuit (IC) layouts, and the more general ImageNet [15] image dataset. We found that the DNN reconstructions attain better visual quality for IC layouts at low photon counts (one$\sim$two per pixel per frame) than for ImageNet.

DNNs represent a very versatile method for inferring the relationship between objects and their corresponding measurements through the imaging system. A DNN is typically trained on a set of examples, each example containing the ideal image of the object (the ground truth) and a corresponding measurement. The DNN can be viewed as an operator mapping the measurement (or a known function of the measurement) to the desired image. The internal parameters of the DNN are adjusted to minimize a loss function that describes how close the image is to the ground truth. After the training, examples from a test set, which have not been used in the training phase, are given to the DNN, which then outputs the reconstructed images.

The phase retrieval problem addressed in this work can be written, for an optically thin object, as:

$$\mathbf{g}(x,y) = \left| F_L \left[u_{\text{inc}}(x,y)\mathbf{t}(x,y)e^{\jmath \mathbf{f}(x,y)} \right] \right|^2, \qquad (1)$$

where (x, y) are the lateral coordinates, $\mathbf{g}$ is the intensity measurement in the detector plane, $\mathbf{t}$ and $\mathbf{f}$ are, respectively, the modulus and phase of the field immediately after the object, u_{inc} the incident field in the object plane, and F_L the Fresnel propagation operator over a distance L. In what follows, we assume that the object modulates only the phase, therefore $\mathbf{t}(x, y) = 1$, and we define: $\mathbf{g} = H(\mathbf{f})$. The optimization problem implicitly solved by the DNN can be written as:

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\text{argmin}} \;\; \psi \left\{ H(\mathbf{f}), \mathbf{g}, \mathbf{\Theta}(\mathbf{f}) \right\}, \qquad (2)$$

where ψ is the functional to minimize and $\mathbf{\Theta}$ the regularizer operating on $\mathbf{f}$, *i.e.* imposing constraints on the solution. In a classical optimization procedure, the regularizer would be chosen *ad hoc*. Instead, here we let the DNN discover a regularization adapted to the specific class of objects we train with.

In this work, the loss function to be minimized is chosen as the negative Pearson correlation coefficient (NPCC) defined in the supplementary material. The use of the NPCC as a loss function, as opposed for example to the mean square error, proved to be a better metric for DNN training in the context of phase retrieval, especially with sparse objects [5].

For our phase retrieval problem, one possibility is to train the DNN with $(\mathbf{f}_k, \mathbf{g}_k)$ couples, k being the index within the training set. We refer to this approach as the "end-to-end" method as it makes use of the endpoints of the optical system, *i.e.* the object phase $\mathbf{f}$ and raw intensity measurement $\mathbf{g}$. It should be noted that, in the end-to-end method, in addition to the regularization, the DNN carries the burden of learning the law of Fresnel propagation. Since Fresnel propagation is a well characterized physical law, it seems inefficient to have the DNN being optimized, even partially, to explain it. Some knowledge about the physical laws has to be included in the training process in order for the DNN to focus on learning a regularizer.

The phase retrieval problem described in Eq. 1 cannot be inverted directly, simply because the detector is not sensitive to phase. Therefore, there is no unique way of disentangling the contribution of the physics and the contribution of the noise (or any other stochastic process involved). However, the well-known Gerchberg-Saxton (GS) [16] and the gradient descent algorithms for phase retrieval provide a useful insight. Even though the phase is not known in the detector plane, an approximate phase can be assumed and used to project the field back to the object plane using the inverse Fresnel operator. In this work, we associate the phase of the incident beam in the detector plane with the square root of the intensity measurement to produce a complex field, which is propagated back to the object plane. The phase of this complex field in the object plane is referred to as an "approximant" (or GS-approximant as it is inspired by the GS algorithm) as it is generally closer to the solution than the raw intensity measurement. Note that the adjoint of operator H, used in the gradient descent method, can also be used to generate an approximant, however, we will restrict our analysis to the GS-approximant. The approximant can be used *in lieu* of the raw measurement for the DNN training. This is an example of a "physics-informed" method as part of the physical process is embedded in the approximant itself. A similar procedure involving such a preprocessing step has been described recently in [17].

In what follows, we describe a series of experiments designed to systematically compare the end-to-end, physics-informed (using the GS-approximant), and the classical Gerchberg-Saxton methods for different levels of noise. Corresponding simulations have been performed and are presented in the supplementary material. The experimental apparatus is depicted in Fig. 1.

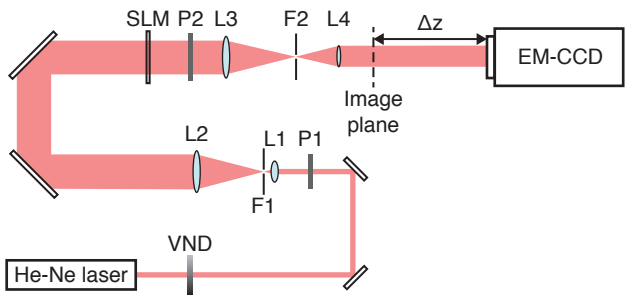

FIG. 1. Optical apparatus. VND: variable neutral density filter, P1-P2: polarizers, L1: 10×, 0.25 NA objective, L2: 100 mm lens, L3: 230 mm lens, L4: 100 mm lens, F1: 5 µm pinhole, F2: iris. SLM: transmissive spatial light modulator. Lens L3 and L4 are confocal. The distance between the SLM and L3 is 230mm, the distance between L4 and the image plane is 100mm, and $\Delta z = 400$mm.

The light source is a Helium-Neon laser emitting continuous wave radiation at 632.8 nm. The laser beam intensity is controlled by a calibrated variable neutral density filter. The beam is focused onto a 5 µm circular pinhole using a 10×, 0.25NA Newport objective. After the pinhole, the beam is collimated with a 100mm lens. The beam is then passed through a transmissive spatial light modulator (SLM) (Holoeye LC2012) with 36 µm square pixels. In order to maximize the SLM phase modulation capability, the incident light is linearly polarized (P1) at a certain angle (45° from the horizontal axis). The modulated light from the SLM is filtered by a second polarizer (P2). The complex (phase and intensity) transmittance of the SLM was calibrated interferometrically for the particular polarizers configuration used in the experiment. The SLM surface is reduced by a factor of 2.3 by a telescope system (lenses L3 and L4 in Fig. 1) in order for the diffracted pattern to fit within the detector. The detector is an EM-CCD 1004×1002 array (QImaging Rolera EM-C2) of 8×8 µm pixels. The EM gain and exposure time of the camera are controlled by software. The detector is placed at a distance $\Delta z = 400$ mm from the image plane. An additional neutral density filter with an optical density of 2 is placed in front of the detector to suppress background light and adjust the photon level range. The actual optical power is measured between filter F2 and lens L4 with a Silicon detector. Details about the calibration are given in the supplementary material. It should be noted that the SLM has a residual intensity modulation effect, which was measured during the calibration step (see supplementary material).

For each image category (ImageNet and IC layouts)

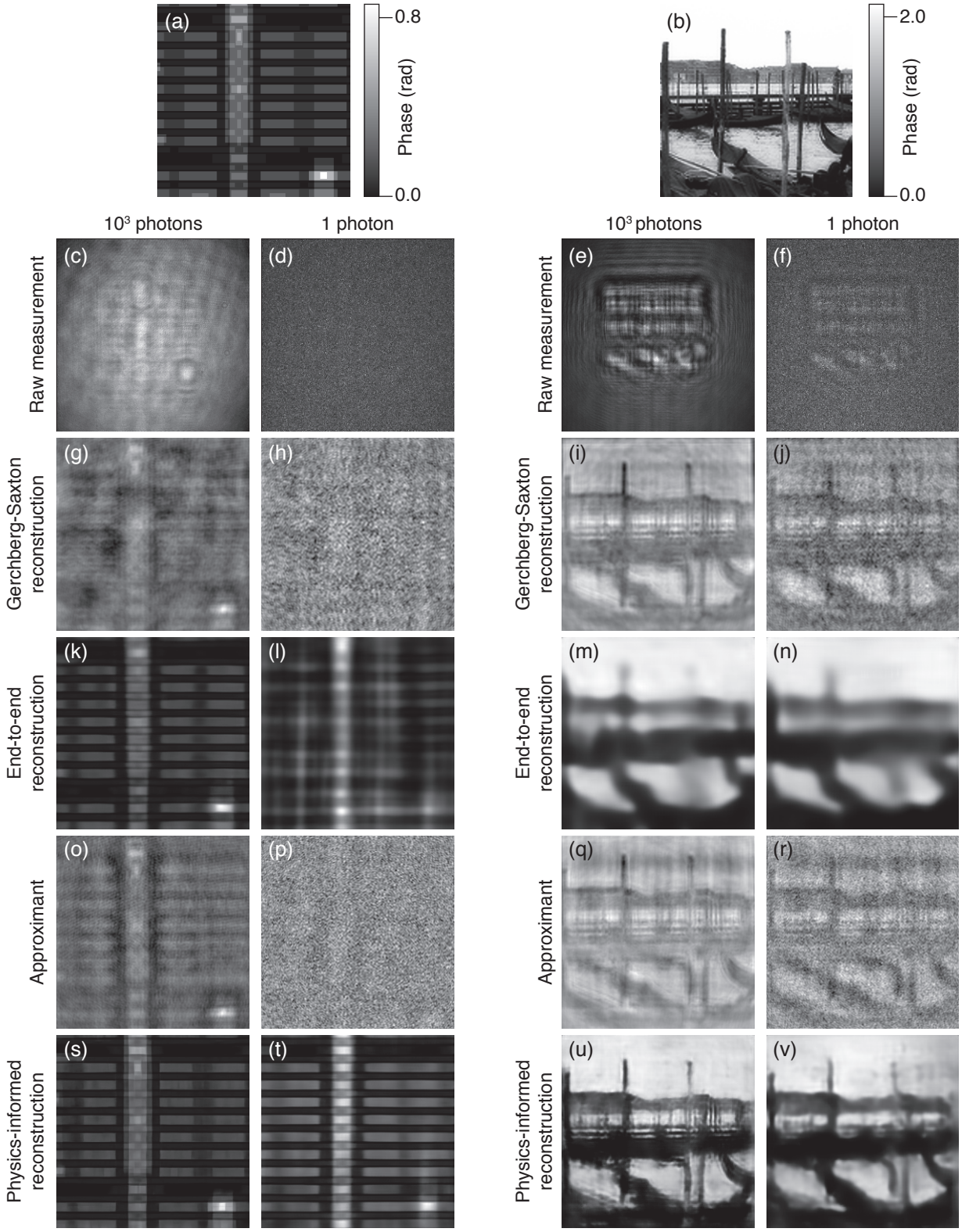

FIG. 2. (a-b) Ground truth phase of one example from each test set of IC layouts and ImageNet. (c-f) Raw measurements in the detector plane. (g-j) Gerchberg-Saxton algorithm reconstructions from the raw measurements c-f. (k-n) DNN reconstructions with the end-to-end method. (o-r) Approximants in the image plane. (s-v) DNN reconstructions from the approximants o-r with the physics-informed method. For better display, the grayscales of all images have been normalized to range from the minimal to the maximal value. Images a, b and g to v represent a phase in the image plane and have a physical size of 4×4mm, while images c to f represent an intensity in the detector plane and have a physical size of 8×8mm.

and for each noise level, a different DNN is trained. The examples are split into a training set, a validation set and a test set containing 9,500, 450 and 50 examples of 256 by 256 pixel images, respectively. The DNN input and output images are 256 by 256 pixels, which is the native resolution of the images in the dataset. For the end-to-end method, the detector images (1002 by 1002 pixels) are resampled to the proper size using bilinear interpolation. For the physics-informed method, each detector image is zero-padded to a size such that the inverse Fresnel propagator would yield an approximant in which the object covered a 256 by 256 pixel area. The DNN has the same encoder-decoder architecture as presented in [5] except that five instead of six convolutional layers are used in the encoder and decoder parts.

Examples of reconstruction from the test sets for both ImageNet and IC layouts are shown in Fig. 2 for two extreme photon level cases. Table I summarizes the noise level for each experiment shown in Fig. 2 and 3. The noise levels indicated in the table refer to the incident beam, *i.e.* with no modulation on the SLM. When a pattern is displayed on the SLM, the SNR at the detector plane varies strongly spatially as a result of intensity redistribution, which is why using the incident beam as reference was preferred. The integration time was set at 2 ms for all experiments mentioned in Table I and Fig. 2 and 3. The integration time was kept short to avoid degradation of the SNR due to air turbulence.

The results shown in Fig. 2 allow us to draw qualitative conclusions. As can be seen in Fig. 2 (g-j) and (o-r), the DNN is very efficient in suppressing the granularity typical of shot noise. The end-to-end method reconstructions appear as low-pass filtered versions of the original image, especially for ImageNet examples. IC layout examples are still reconstructed with sharp edges as this feature is omnipresent in the IC layout. The interpretation is that the DNN does not fully learn the diffraction operator, but rather learns how to suppress fringes and other diffraction related patterns and also how to promote characteristic features of the training examples. The physics-informed reconstructions are visually better because, in this case, high frequencies are provided to the DNN by the approximant (especially visible in Fig. 2q). In the low photon example of the IC layout (Fig. 2t), the general pattern is recovered, but additional spurious tracks have been added by the DNN that seems to promote periodicity, a feature quite prominent in IC layout examples.

TABLE I. Noise levels and photon count for the experiments shown in Fig. 3. The illumination conditions are the same for both the IC layout and the ImageNet datasets. The photon count is the effective number of photons after dividing by the quantum efficiency per detector pixel averaged over the whole detector field for the incident beam (no modulation on the SLM). The procedure for measuring the photon count is given in the supplementary material. The SNR is the mean of the incident beam signal divided by its standard deviation and averaged over the whole field of view. The limit SNR is the square root of the number of photons.

Experiment	EM gain	Photon count ±5%	SNR	Limit SNR
1	1	1.0×10^3	20	32
2	1	84	2.7	9.2
3	1	43	1.45	6.6
4	4.8	9.8	0.9	3.1
5	54	1.1	0.5	1.0
6	54	0.25	0.24	0.5

We use the Pearson correlation coefficient (PCC = −NPCC) as a figure of merit for the quality of the reconstructions; the results are shown in Fig. 3. Note that other metrics for image quality can be used. In Fig. 3 of the supplementary material, we show a comparison of the following metrics: the classical Mean Square Error (MSE), the Structural Similarity Index (SSIM) [18], and a wavelet transform based SSIM [19].

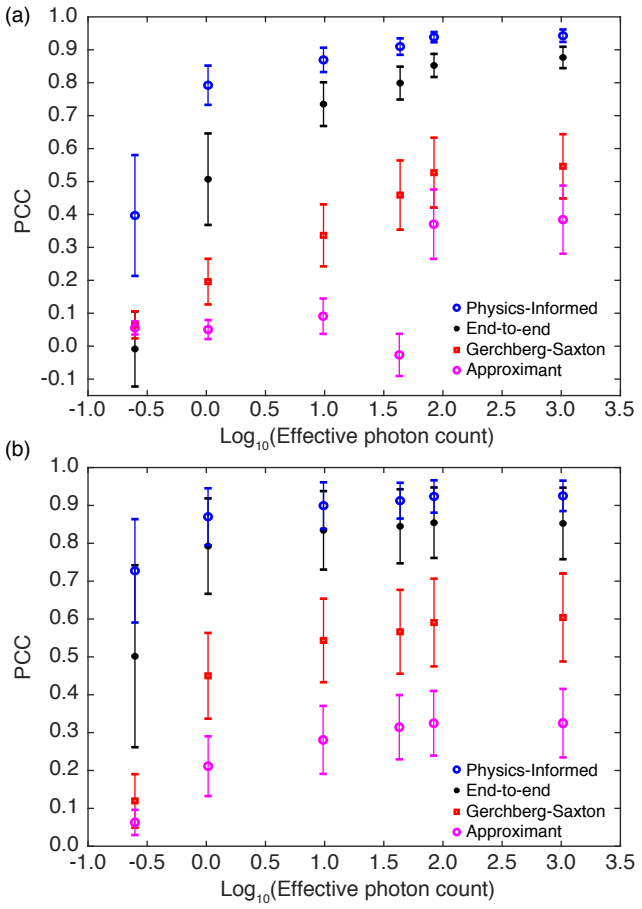

FIG. 3. Pearson correlation coefficient between the ground truth and the DNN reconstructions. (a) IC layout data set. (b) ImageNet data set. The markers indicate the mean over the test set (50 examples) and the error bars ±1 standard deviation from the mean.

In the case of the IC layout, for all photon levels, the

physics-informed method performs systematically better than the end-to-end method, which in turn performs better than the GS algorithm. A similar result holds for the ImageNet example set, except that there is less difference between the end-to-end and the physics-informed reconstruction and also that the standard deviation of the reconstruction quality is larger even for high photon levels. The GS reconstructions for high photon level do not display this trend (their standard deviation remains equally large). This latter observation confirms that the strong prior in the IC layout geometry is efficiently exploited by the DNN. In Fig. 3, we also plotted the PCC between the ground truth and the approximant. The approximant is the input image to the physics-informed DNN and is also the result of the first iteration of the GS algorithm. As such, the increase in image quality between the approximant and the GS and physics-informed reconstructions indicates the improvement brought by each technique. The improvement brought by the DNN is systematically better. We did not plot the PCC between the raw measurement and the ground truth as these images belong to different spaces (object vs. detector space), and for strong diffraction the comparison would be meaningless.

The PCC is not sensitive to the magnitude of the images (*i.e.* $\mathrm{PCC}(A,B) = \mathrm{PCC}(\alpha A, \beta B), \alpha, \beta \in \mathbb{R}$), the phase images are thus reconstructed up to a scaling factor. However, for a given DNN the scaling factor is constant and can be retrieved by comparing the validation set ground truth examples and corresponding reconstructions. In practice, the scaling factor is obtained by comapring the histograms of the ground truths and reconstructions images.

The approximant clearly helps in recovering high fidelity images. The question of knowing what is the best way of obtaining an approximant in the context of phase retrieval is beyond the scope of this paper. It should be recognized that the GS-approximant the way it is computed here corresponds to half of the first iteration of the GS algorithm. The question whether it is worthy to iterate more in order to generate an approximant is still open, but preliminary results tend to show that little is gained by iterating more.

This work was supported by the Intelligence Advanced Research Projects Activity (IARPA) FA8650-17-C-9113.

* agoy@mit.edu
† Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore 117543, Singapore.

[1] L. Xu, J. S. Ren, C. Liu, and J. Jia, in *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc., 2014) pp. 1790–1798.
[2] M. Mardani, H. Monajemi, V. Papyan, S. Vasanawala, D. L. Donoho, and J. M. Pauly, CoRR (2017), arXiv:1711.10046.
[3] M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, Scientific Reportsvolume **7**, 17865 (2017).
[4] A. Sinha, J. Lee, S. Li, and G. Barbastathis, Optica **4**, 1117 (2017).
[5] S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, Optica **5**, 803 (2018).
[6] Y. Rivenson, Y. Zhang, H. Gnaydn, D. Teng, , and A. Ozcan, Light: Science and Applications **7**, 17141 (2018).
[7] C. A. Metzler, P. Schniter, A. Veeraraghavan, and R. G. Baraniuk, Archive (2018), arXiv:1803.00212.
[8] Z. D. C. Kemp, Journal of Optics **20**, 045606 (2018).
[9] L. Boominathan, M. Maniparambil, H. Gupta, R. Baburajan, and K. Mitra, CoRR (2018), arXiv:1805.03593.
[10] R. Horstmeyer, R. Y. Chen, B. Kappes, and B. Judkewitz, CoRR (2017), arXiv:1709.07223.
[11] D. G. Sandler, T. K. Barrett, D. A. Palmer, R. Q. Fugate, and W. J. Wild, Nature **351**, 300 (1991).
[12] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis, Optica **2**, 517 (2015).
[13] U. S. Kamilov, I. N. Papadopoulos, M. H. Shoreh, A. Goy, C. Vonesch, M. Unser, and D. Psaltis, IEEE Transactions on Computational Imaging **2**, 59 (2016).
[14] C. Chen, Q. Chen, J. Xu, and V. Koltun, CoRR (2018), arXiv:1805.01934.
[15] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009) pp. 248–255.
[16] R. W. Gerchberg and W. O. Saxton, Optik **35**, 237 (1971).
[17] Y. Sun, Z. Xia, and U. S. Kamilov, Opt. Exp. **26**, 14678 (2018).
[18] Z. Wang, E. Simoncelli, and A. Bovik, in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003* (2003) pp. 1398–1402.
[19] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, IEEE Trans. on Im. Proc. **18**, 2385 (2009).