PHOTONICS Research

Toward simple, generalizable neural networks with universal training for low-SWaP hybrid vision

Baurzhan Muminov, Altai Perry, Rakib Hyder, M. Salman Asif, and Luat T. Vuong*

University of California Riverside, Riverside, California 92521, USA
**Corresponding author: LuatV@UCR.edu*

Received 7 December 2020; revised 9 March 2021; accepted 25 March 2021; posted 19 April 2021 (Doc. ID 416614); published 14 June 2021

Speed, generalizability, and robustness are fundamental issues for building lightweight computational cameras. Here we demonstrate generalizable image reconstruction with the simplest of hybrid machine vision systems: linear optical preprocessors combined with no-hidden-layer, "small-brain" neural networks. Surprisingly, such simple neural networks are capable of learning the image reconstruction from a range of coded diffraction patterns using two masks. We investigate the possibility of generalized or "universal training" with these small brains. Neural networks trained with sinusoidal or random patterns uniformly distribute errors around a reconstructed image, whereas models trained with a combination of sharp and curved shapes (the phase pattern of optical vortices) reconstruct edges more boldly. We illustrate variable convergence of these simple neural networks and relate learnability of an image to its singular value decomposition entropy of the image. We also provide heuristic experimental results. With thresholding, we achieve robust reconstruction of various disjoint datasets. Our work is favorable for future real-time low size, weight, and power hybrid vision: we reconstruct images on a 15 W laptop CPU with 15,000 frames per second: faster by a factor of 3 than previously reported results and 3 orders of magnitude faster than convolutional neural networks.

https://doi.org/10.1364/PRJ.416614

1. INTRODUCTION

Image reconstruction has wide application in medicine [1,2], biology [3], X-ray crystallography [4], and low-light vision, among other technologies. These reconstructions generally involve solving an inverse problem and retrieving the phase from phaseless intensity measurements. The field has been an active area of research for several decades [5–7] and inverse solvers achieve impressive results with additional coded optics or optical scanning [8–21]. More recently, deep neural networks, and specifically convolutional neural networks, enable single feed-forward, noniterative reconstruction [22] and are capable of learning from the statistical information contained in a variety of systems, from speckle [23,24] to coded diffraction [25] patterns. Inverse solvers using neural networks are generally faster than iterative, optimization-based, or optical scanning-based algorithms and may require as few as 100 illumination training patterns, for example, with an "unrolled" neural network [26].

However, despite the benefits of using neural networks to solve inverse problems, there are also drawbacks. Some of these issues—especially those associated with phase retrieval—have been solved. Others—particularly those related to generalizability, robustness, and processing time or energy—remain active areas of research [24]. Since neural networks learn how to weigh the importance of information patterns based on training data, they exhibit a tendency to "memorize" patterns to gain intuition about the task [27]. This predisposition toward prior data is advantageous for building "inductive, artificial intelligence machines" that extract patterns; however, that predisposition is a detriment to the generalizability of inverse solutions, e.g., for building real-time computational cameras. Antun *et al.* [28] highlight three specific issues encountered by neural nets in imaging tasks:

1. Small, sometimes undetectable perturbations in the input (both image and sampling domain) can cause severe artifacts in the image reconstruction.
2. Small structural changes can be left undetected.
3. More samples in the training set can lead to a deterioration of the results (as a result of the "memory" effect described above). Subsequently, algorithms themselves can stall or experience instabilities.

Whereas biomedical applications are aimed at large-image, high-quality image reconstruction [3], we turn our attention toward building real-time computational cameras for low size, weight, and power (SWaP) image reconstruction, which are needed for autonomous-vehicle applications. In our prior effort [29], we demonstrate reconstruction with a "small brain" dual-layer neural network. Such regression-based approaches [30] demonstrate fast reconstruction rates, robustness to noise, and show potential for generalization with a phase vortex

encoder. Here, we focus entirely on the generalizability of a simple neural network using a single-layer architecture for image reconstruction. We supply the model with a generalized or universal training set (UTS) (synthetic images, used to train the neural network) and then test the neural network with images of different, unseen classes [see Fig. 1(a)]. A UTS-trained model overcomes the challenges associated with the "stereotypes" that generally arise from training by a specific image set. On the other hand, some disadvantages include the fact that the neural network is too simple to reconstruct images when nonlinear transformations are required [31]. Nevertheless, our results provide insight for training generalizable neural networks and computational cameras that operate at fast speeds. Our proposed method can readily be used for the initialization of alternating minimization problems or downstream image analysis tasks [32–34].

It is perhaps surprising that the simple learning model possesses enough capacity to recover a good approximation of the inverse coded-diffraction problem, and even with such a simple neural network there are interesting issues to address. In an effort to move toward producing a generalized training set, we compare the performance of the vortex encoder with other random encoders. From there, we build intuition for the UTS design based on the modal decomposition of the training, diffracted imaging patterns, and singular value decomposition (SVD)-entropy. We also perform experiments, which build heuristics for real-world applications. We find that the choice of training images and optical encoder is important for achieving generalizability, since not all imaged patterns provide a unique mapping to be learned and not all learned intensity patterns aid image reconstruction. While we have not quantitatively analyzed the image reconstruction, i.e., compared the set of training images to the span of the neural network, we observe that reduced SVD-entropy in the training set increases the learning efficiency, in both simulations and experiments.

2. PROJECT SETUP

In this section, we review an approach similar to Ref. [29] for our study of generalizable training. Figure 1(b) shows a schematic of the hybrid machine vision system, which encodes the image prior to the neural network with either a random or vortex phase pattern.

A. Hybrid Vision System

The fields from the object at the diffractive encoder plane are $F(x,y)$. The encoder plane is imprinted with two diffractive element patterns $M(x,y)$, as shown in Fig. 1(c). A sensor or detector captures the intensity pattern of the electric fields $F'(u,v)$. Let $\mathcal{F}$ be the Fourier transform operation $(x,y) \rightarrow (u,v)$, where we capture an image in the Fourier plane:

$$F'(u,v) = \mathcal{F}[M(x,y)F(x,y)]. \tag{1}$$

Light from each object produces two images, each with a different diffractive element $M(x,y)$. Although the mask pattern may imprint vector (i.e., polarization-dependent) or spectral (time-dependent) delays, here we assume a homogeneous polarization, a linear encoder, and monochromatic,

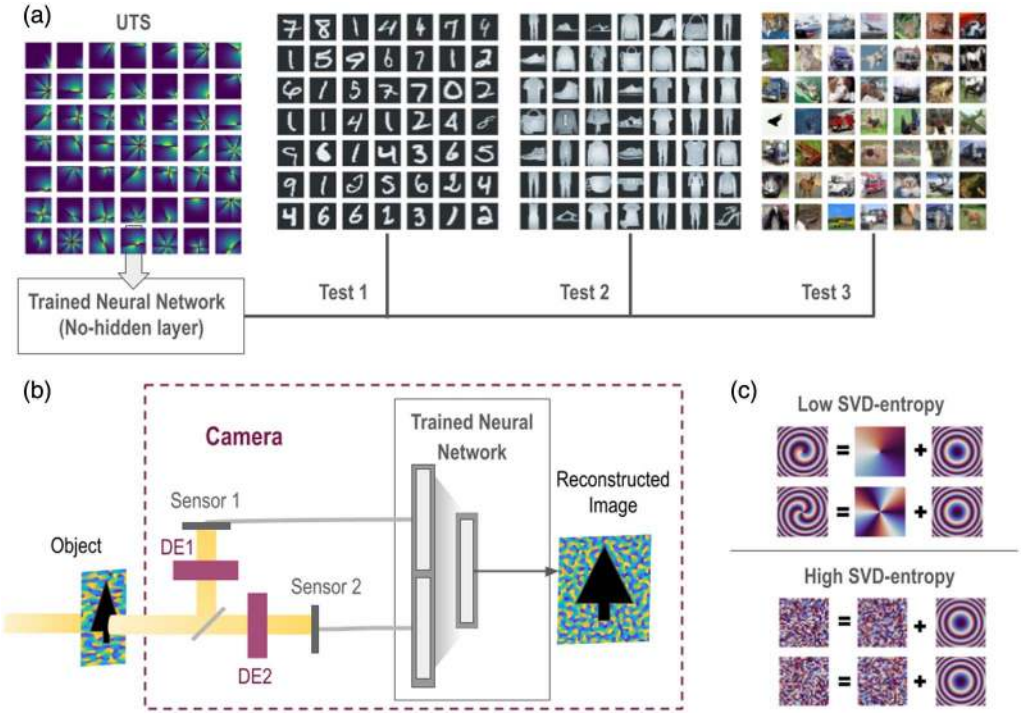

Fig. 1. (a) Project objective: design a generalized training set for a neural network, which can later be used for general image reconstruction without retraining and can operate in real time. (b) Schematic of hybrid vision camera where light from an object is transmitted through a diffractive encoder (DE). Sensors capture two transmitted images that are combined as inputs to the trained neural network, which reconstruct the object from the detector-plane images. (c) This project employs two pairs of diffractive encoders: one with low SVD-entropy (lens and topological charge $m = 1$ and 3) and the other with high SVD-entropy (uniformly distributed random pattern).

continuous-wave light. All optical neural networks have been previously demonstrated, notably with several diffractive layers in the terahertz regime [35], with nonlinear activations via saturable-absorbing nonlinearities [36], and with nano-interferometric etalons [37] in the visible regime. All-optical methods maximize speed and minimize energy loss in the neural computation [38]. At the same time, all-optical systems require nonlinear interactions as proxies for the electronic neural network layer activations. These nonlinearities occur at small length scales in order to confine light sufficiently, so all-optical computing may be more sensitive to environmental conditions and less suitable for autonomous-vehicle computational cameras.

By contrast, we focus on hybrid imaging in which optical processing conditions sensor measurements and an electronic neural network performs reconstruction [39–41]. Our work is also inspired by pytchography approaches in Refs. [10–12]. Two phase masks are used to capture the intensity measurements of the object on the sensor, which are then fed to a no-hidden-layer neural network. At this time, we do not predict depth sensing with imaging, so the masks contain lenses for Fourier-plane detection. Here we reproduce the object based on the detector intensity patterns and assume that the detector is in the focal plane associated with a quadratic radial phase of the mask. In recent work, Fresnel mid-field imaging shows potential for better object-based depth detection [42].

In a manner similar to Ref. [29], we generate phase-modulated patterns:

$$F(x,y)M(x,y) = e^{i\alpha \boldsymbol{X}} G(x,y)M(x,y), \tag{2}$$

where $G(x,y)$ is the Gaussian beam pattern illuminating the object and $\boldsymbol{X}$ is the positively valued original image. This Gaussian pattern represents a smooth pupil function or the illuminating beam. In our study, we fix $\alpha = \pi$ and find that the reconstruction quality does not change significantly when α varies from $\pi/4$ to $3\pi/2$.

The general inverse problem for mapping the detector measurements to the original image involves solving the following nonlinear system of equations:

$$\boldsymbol{Y} = H(\boldsymbol{X}) + \boldsymbol{N}, \tag{3}$$

or for our specific case,

$$\boldsymbol{Y} = |\mathcal{F}[e^{i\alpha \boldsymbol{X}} G(x,y)M(x,y)]|^2 + \boldsymbol{N}, \tag{4}$$

where $\boldsymbol{Y}$ is the positively valued sensor measurement; $H(\cdot)$ is a nonlinear transform operator that includes the transfer function of the optics, light scattering, and the sensitivity curve of the detector; and $\boldsymbol{N}$ is the measurement noise.

The Fourier-plane intensity patterns $\boldsymbol{Y}$ are the inputs to a neural network. The neural network estimates $\boldsymbol{X}$ (size 28 × 28) given $\boldsymbol{Y}$ (size 28 × 28 × 2). To train the neural network, we use the TensorFlow library with the mean squared error loss and Adam optimization algorithm. Convergence is achieved with similar results using either "linear" or "ReLu" activation. Our approach is simple and shows promising opportunities for generalized image reconstruction with "small brain" neural networks.

B. Universal Training Sets and Diffractive Encoders

We choose two pairs of diffractive encoders. One pair is composed of vortex masks, where each mask has an on-axis singularity of either $m = 1$ or 3:

$$M(x,y) = e^{-(x^2+y^2)\left(\frac{i}{f\lambda}+\frac{1}{w^2}\right)} e^{im\phi}, \tag{5}$$

where f is the effective focal length of the radial quadratic phase, λ is the wavelength of light, m is an on-axis topological charge, and w is the width of the Gaussian beam illuminating the mask. Figures 1(b) and 1(c) show diffractive elements with m = 1 and 3. The second pair is composed of random masks, where each pixel of the transmitted pattern is encoded with a random phase from 0 to 2π. The mask is also illuminated with the same Gaussian beam. On the side of the training, we work with a range of images composed of 28 × 28 patterns that are random $\boldsymbol{X}_R$, Fourier-based $\boldsymbol{X}_F$, or shapes related to a vortex phase $\boldsymbol{X}_V$.

We approach the generalized training to understand the modal distribution of each image $\boldsymbol{X}$. In principle, the training images should span the space of the test images, which defines the requirements for reconstruction. This would suggest that each coded-diffraction Fourier-plane image should be decomposed into Fourier modes, since this common basis provides a unique and straightforward basis for each image. Such Fourier patterns are linear wave patterns that change with phase and vary with variables j, k, l, n:

$$\boldsymbol{X}_{F(s_j,s_k,\phi_l,n)}(x,y) = \angle[e^{i(xs_j+ys_k+\phi_l)}]G_n, \tag{6}$$

where combinations of $s_j = 2\pi j/\mathrm{d}x$, $s_k = 2\pi k/\mathrm{d}y$, and k span the Fourier space intended to reproduce any arbitrary image and N. G_n represents a scanning Gaussian beam with varied width and center,

$$G_n(x,y) = e^{-[(x-x_n)^2+(y-y_n)^2]/w_n^2}, \tag{7}$$

where x_n, y_n, w_n tune size of the UTS to be comparable to others. The size of the dataset also changes the phase shift, where $\phi_k = 2\pi k/N$ and N is the number of the uniquely valued wave fringes with wavenumbers s_j, s_k in $\boldsymbol{X}_F$.

We refer to a "vortex training set" as a UTS composed of shapes similar to the phases of a vortex beam that have distinct edges and curves:

$$\boldsymbol{X}_{V(x_j,y_k,\phi_j,n,l)}(x,y) = \angle\{e^{im_l \arctan[(y-y_k)/(x-x_j)]+\phi_k}\}G_{j,k,n}. \tag{8}$$

For the vortex $\boldsymbol{X}_V$ as well as the random $\boldsymbol{X}_R$ UTS, we use uniformly distributed random variables to mask the pattern with a Gaussian profile. In other words, combinations of x_j, y_j, and $\phi_k = 2\pi k/N$ span the dataset, or

$$G_{j,k,n}(x,y) = e^{-[(x-x_j)^2+(y-y_k)^2]/w_n^2}. \tag{9}$$

This Gaussian function $G_{j,k,n}(x,y)$ represents a scanning light beam that illuminates the training images. All image patterns are positively valued and normalized to have a peak value of 1.

We produce three UTSs that span the image space using up to 40,000 patterns. The goal of our project is to illustrate trends and intuition with these datasets.

Once trained with a large dataset, we observe that the dense neural network without hidden layers can approximate almost

any shape-based image (MNIST, fashion MNIST, CIFAR). An example set of reconstructed images from different classes is shown in Fig. 2. Figure 2 shows a representative set of images reconstructed from models trained with $\boldsymbol{X}_F$, $\boldsymbol{X}_V$, and $\boldsymbol{X}_R$ and a vortex mask. In each case, 20,000 training images are used. Error with thresholding is as low as 10% with test datasets. While the overall error is similar, models trained with the vortex-phase datasets, $\boldsymbol{X}_V$, generally have the lowest error and strongly highlighted edges. Meanwhile models trained with a Fourier basis $\boldsymbol{X}_F$ have the highest error and models trained with a random basis $\boldsymbol{X}_R$ have error in between, with error distributed over the area of the image. Additional differences are explained in the following section.

C. Differences in Convergence and Single-Pixel Response with Different Training Sets

With this simple neural network and three different UTSs, we observe trends in convergence and overfitting. These trends consistently depend on the choice of the UTS patterns regardless of the choice of mask M_V or M_R. Figures 3(a)–3(c) show samples from 20,000-image $\boldsymbol{X}_F$, $\boldsymbol{X}_V$, and $\boldsymbol{X}_R$ UTS with the vortex mask M_V. Some pairings converge with minimal overfitting while others do not provide enough information in $\boldsymbol{Y}$ to calculate the inverse of the nonlinear mapping, $H(\boldsymbol{X})$ [Figs. 3(d)–3(f)].

A Fourier basis is the most well-known spectral basis for decomposing an image. When training with a Fourier basis, the validation loss stops decreasing after a certain number of epochs, which signals that the neural network struggles to extract information about the mapping given this orthogonal set of images. What this tells us is rather unintuitive about the span or basis of image reconstruction with neural networks, but potentially addressed in Ref. [43]: the images are less effectively learned by the neural net because there is minimal overlap between them; the correlations between Fourier modes are less visible to the neural net.

The random UTS also unreliably converges when the dataset is smaller than 2000, and its loss generally shows a "hill,"

Fig. 2. Reconstructed images from (a), (b), (c) MNIST handwritten and (d), (e), (f) fashion MNIST datasets with random, Fourier, and vortex bases, respectively. The vortex basis provides edge enhancement for object detection. (g) Ground truth and (h) reconstructed images from the CIFAR-10 dataset using the vortex training bases and a vortex mask as the encoder.

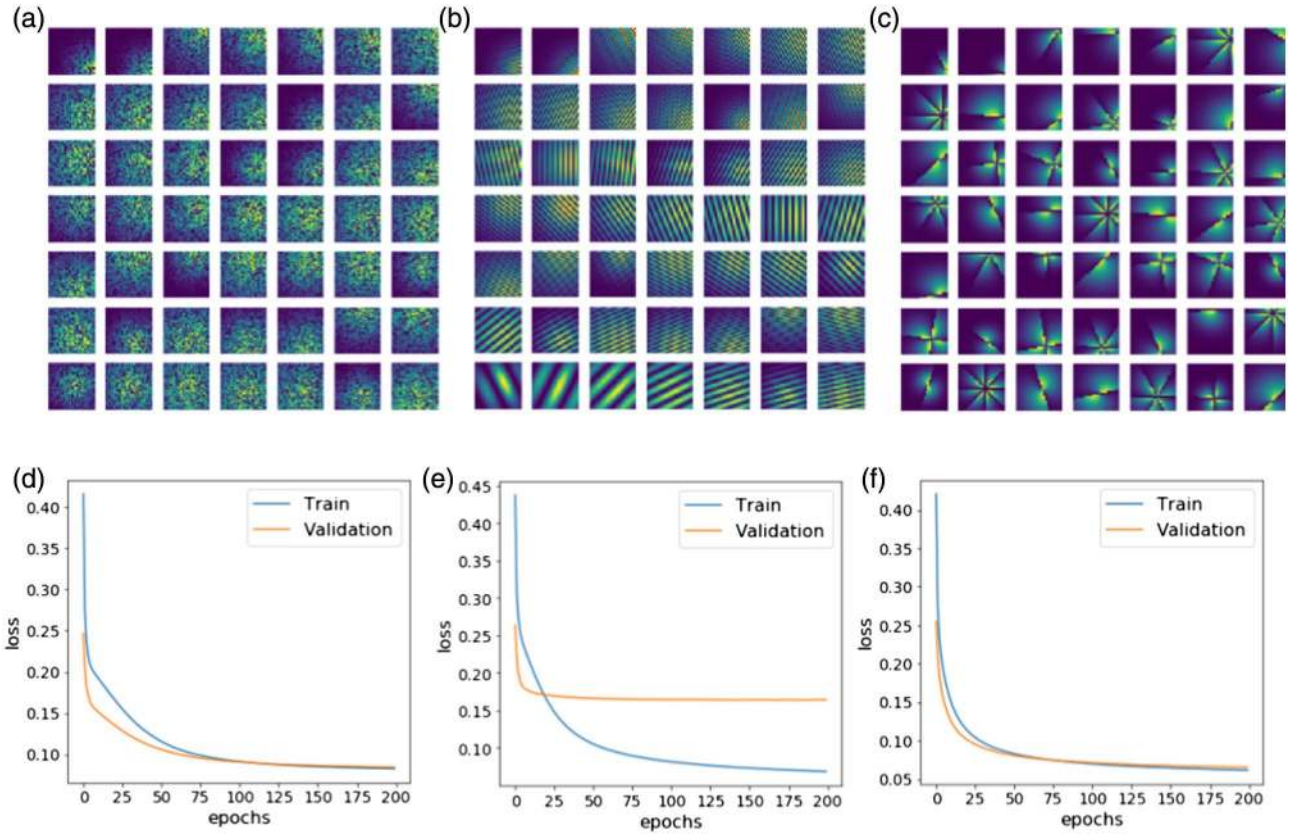

Fig. 3. (a)–(c) Sample training images $\boldsymbol{X}_R$, $\boldsymbol{X}_F$, and $\boldsymbol{X}_V$ or random, Fourier, and vortex training sets. (d)–(f) Corresponding training and validation curves.

where the loss plateaus before dropping. Meanwhile, the vortex-based UTS is less prone to such behavior. This combination of trends tells us that neither orthogonality nor randomness is ideal for training a neural network. The structured pattern of our vortex-based UTS $\boldsymbol{X}_V$ is a better candidate for generalized training compared to random $\boldsymbol{X}_R$ or Fourier $\boldsymbol{X}_F$ patterns. In our discussion, we provide some measures related to the UTS image analysis and trained model robustness.

3. DISCUSSION

In this section, we discuss the ability to recreate sharp images, which may be seen by the single-pixel response. The single-pixel response from the random UTS-trained neural network is sharply corrugated [Fig. 4(a)], whereas the structured, single-pixel images from the vortex-trained model are generally smooth with a sharp "hole" in the center or dark spot [Fig. 4(b)]. We claim that these differences in the impulse response are responsible for the edge-enhanced reconstruction of shapes in Figs. 2(c) and 2(f). Figures 4(a) and 4(b) illustrate example images reconstructed with just one "hot" pixel in the camera sensor plane. These patterns are the building blocks of the reconstruction scheme and these patterns change depending on how the model is trained. Depending on the training set, the model is tuned to pay attention to different features of the image, which may depend on the task at hand.

Figure 4(c) provides a simple noise analysis that shows the additional advantage of robustness when the neural network is trained with a low-entropy UTS. We show the reconstruction error as a function of noise magnitude. Poisson shot noise and background noise are added to the Fourier-plane intensity patterns of the test image set. Low SVD-entropy image training and encoders appear more robust.

A. Analysis with Singular Value Decomposition Entropy

In order to estimate complexity of the pattern we employ the measure of entropy. We approximate the 2D entropy of the images using the spectra of singular value decomposition (SVD), which describes the complexity of an image. Unlike Shannon entropy [44], SVD-entropy illustrates the mixture of spatial modes that are present in an image.

We use a normalized relation for the SVD-entropy that is invariant with image intensity scaling:

$$E_{\mathrm{SVD}} = -\sum_{1}^{K} \bar{\sigma}_i \log_2(\bar{\sigma}_i), \tag{10}$$

where the argument $\bar{\sigma}_i$ is the normalized magnitude of the singular values or the modal coefficients of the image, given as

$$\bar{\sigma}_i = \frac{\sigma_i}{\sum_1^K \sigma_i} \quad \text{and} \quad \sum_i \bar{\sigma}_i = 1, \tag{11}$$

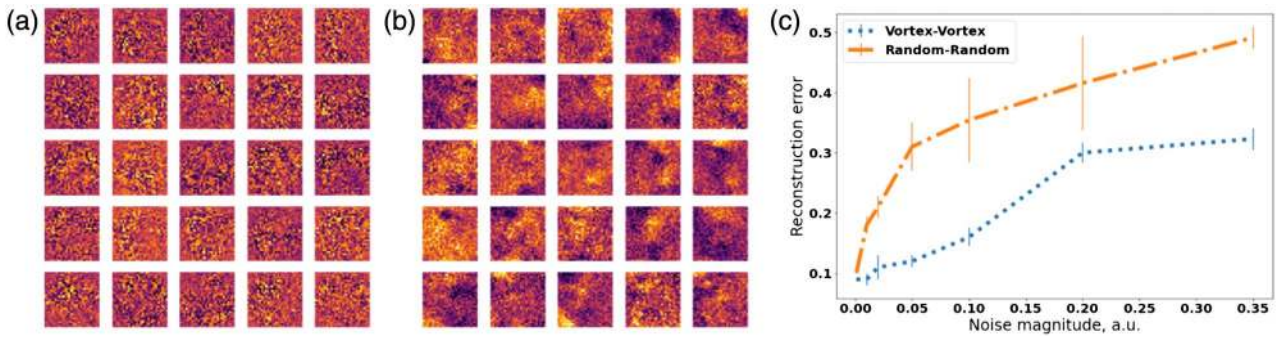

Fig. 4. (a) Single "hot" pixel response of the random model and (b) single-pixel response of the vortex model, which demonstrates sharp edges and resolves high-contrast objects. (c) Comparison of reconstruction error for different levels of noise given high-entropy random UTS and random mask and lower SVD-entropy vortex UTS and vortex mask. This error corresponds to the scenario in which shot noise dominates the background noise.

where K is the number of singular values and σ_i are the singular values.

Some trends related to the SVD-entropy are illustrated in Fig. 5. If images in the set have several high singular values σ_i, the images may be reconstructed using fewer "elementary" patterns; those with higher entropy require many more patterns to achieve enough reconstruction accuracy. Low SVD-entropy images are smoother with fewer edges. On the other hand, images with many discontinuities exhibit a high degree of SVD-entropy.

From our analysis of differently structured patterns, the SVD-entropy scales logarithmically with the edge steps or dislocations in an image [Figs. 5(a) and 5(b)]. In this illustration, we plot the phase of an $m = 3$ vortex with varied Gaussian-beam filtering. The measure of 2D SVD-entropy aids our analysis of the UTS. The vortex UTS has a broad range and lower values of SVD-entropy in contrast to the random UTS [Fig. 5(c)].

Pertaining to our efforts toward generalized training or a UTS, we see that a low SVD-entropy training set like that with structured patterns $\boldsymbol{X}_V$ allows us to extract the structured (low SVD-entropy) information from the data [Figs. 2(c), 2(f), 2(g), and 2(h)]. This effectively acts as a filter for salient features of the image. This low SVD-entropy training would be useful for some specific tasks, especially when, e.g., we are less interested in the image's background information than in the foreground object.

B. Heuristic Experiments

To illustrate the potential and the impact of our approach for generalizable training, we show heuristic experimental results. In simulations, almost any encoded diffraction pattern with a mask presents a learnable map for a simple neural network. However, in practice when noise is present, neural networks do not always converge. Our experimental data show that under noisy experimental conditions where light is unpolarized and the sensor data is collected with significant levels of noise, the high SVD-entropy dataset is not suitable for the task of image reconstruction: background light and distortions render a high SVD-entropy training image useless since the neural network does not learn the pattern. By contrast, a neural network trained on low SVD-entropy images is capable of recovering

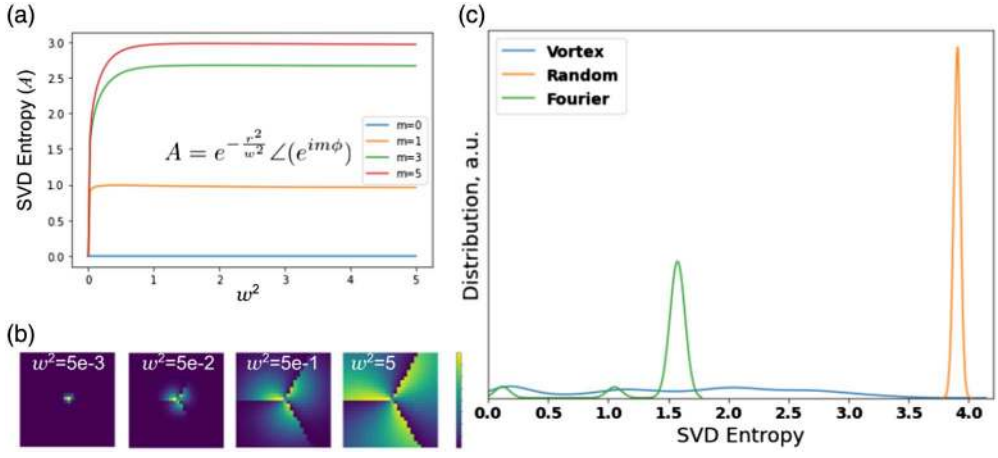

Fig. 5. (a) SVD-entropy of a structured pattern composed of the phase of a vortex (modulus 0, 2π) and a Gaussian mask with radius of w^2. A few-pixel pattern has almost zero entropy, and the SVD-entropy saturates for a vortex depending on the topological charge. (b) Illustration of these patterns with $w^2 = 5 \times 10^{-3}$, 5×10^{-2}, 5×10^{-1}, and 5 corresponding to SVD-entropy values of 0.94, 1.8, 2.6, and 2.7. The SVD-entropy strongly relates to the length of the edge dislocations of an image. (c) Histogram of the SVD-entropy in the vortex $\boldsymbol{X}_V$, Fourier $\boldsymbol{X}_F$, and random $\boldsymbol{X}_R$ generalized training sets implemented in this project.

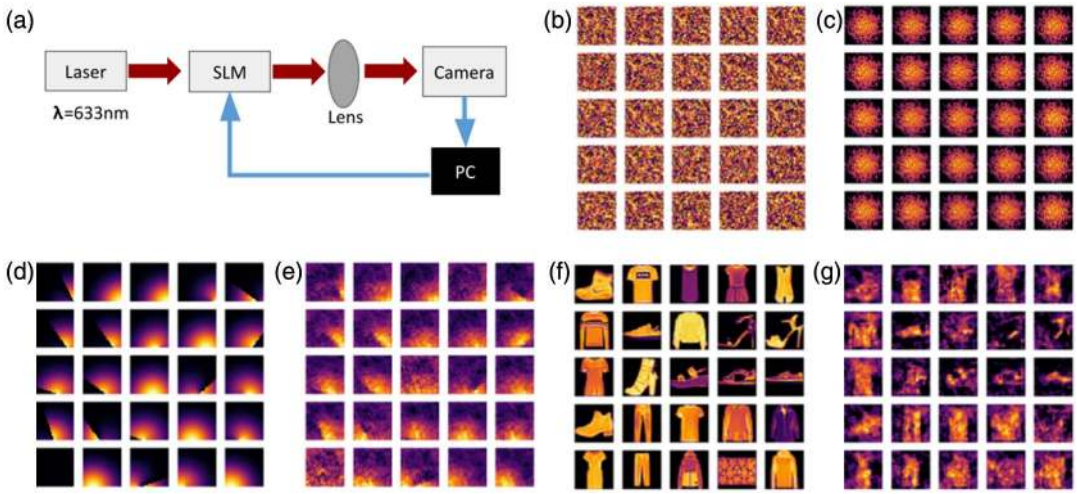

Fig. 6. (a) Schematic of experimental reconstruction with UTS. There is no spatial filter or polarizer, images are noisy, and at this wavelength, the modulation dynamic range is only $\alpha = \pi$. This was done intentionally to simulate poor experimental conditions with background light. (b) Sample of random UTS images and (c) sample of reconstructed images produced by random patterns, which are not learned by the simple neural network model experimentally. On the other hand, (d) simpler images with fewer edges are (e) reconstructed by the neural network. (f) Sample of ground truth images and (g) discernable reconstructed patterns when the neural network is trained by the vortex dataset.

reasonable approximations of the unseen images, as shown in Fig. 6.

Our experimental setup consists of a 633 nm helium–neon continuous-wave laser, microscope objective, HOLOEYE spatial light modulator and focusing lenses, and a CMOS 8-bit camera (1280 × 1024 pixels resolution). The setup does not include polarizers as part of the design to provide a large-background and an unmodulated signal to test the limits of image reconstruction with a simple neural network. As a result, we are unable to recover images with the zeroth-order transmitted pattern. When we instead collect the sensor data at the first diffraction maximum, we are successful with image reconstruction but only with the vortex UTS. For reconstruction purposes, small square patches of the detector pattern are taken (e.g., 50 × 50 pixels).

In our experiments with imperfect spatial beam profiles and background unmodulated noise, the simple neural networks do not converge with random masks [the results are shown in Figs. 6(b) and 6(c)]. Experimentally, we demonstrate two masks shown in Figs. 1(c) and 1(d), which are successfully learned by the neural network. The low SVD-entropy dataset composed of shapes with straight edges and curves, i.e., $\boldsymbol{X}_V$ [Eq. (8)], converges but the high SVD-entropy random $\boldsymbol{X}_R$ patterns do not. Again, we find it more difficult to train a simple neural network with a high SVD-entropy UTS.

4. CONCLUSION

Corners, edges, and higher-order solutions are a challenge in image reconstruction, requiring a higher degree of superposed waves [45]. This more complex representation of images is the definition of SVD-entropy in an image and suggests that the reconstruction of such images requires the learning of images composed of high SVD-entropy patterns [46]. We find, however, that this is not always the case when aiming for robust neural network-based reconstruction. In fact, generalized training with low-entropy patterns recreates these sharp features well with edge enhancement.

We show that a simple neural network without hidden layers is capable of learning generalized image reconstruction. With this simple architecture designed to approach generalized training, it is evident that not all generalized datasets are equal. When we compare the convergence of differently structured datasets such as handwritten digits and fashion MNIST, a set of images or encoder based on vortex phase patterns (structured, low SVD-entropy, a combination of edges and curves) yields image reconstruction with lower error than a high SVD-entropy random encoder pattern that contains many edges. With a dataset such as CIFAR, the salient features are preserved in image reconstruction using a vortex UTS.

We have previously shown that a convolutional neural network can outperform a single-layer neural network but with significantly higher energy cost. The deep neural network is also less robust to noise [29]. Here, we aim to work with a "small brain" neural network rather than a deep neural network architecture. This approach has been specifically tuned with the aim of low-SWaP computational cameras. We draw the following conclusions.

- Single-layer neural networks are capable of approximating the inverse mapping from phaseless Fourier-plane intensity patterns after basic training.
- Such moderate-accuracy generalizable image reconstruction achieves high speeds (we achieve 15,000 frames per second on a 15 W laptop CPU).
- Image reconstruction with simpler neural networks is robust to the vulnerabilities and instabilities described by Ref. [28].
- Even with a simple neural network architecture and a large training basis set, we encounter differences in convergence. (Experimentally with an imperfect encoder, neural networks learn low SVD-entropy images more rapidly and reliably than high SVD-entropy.)
- Low SVD-entropy images are valuable in training neural networks to extract the salient features of the image.

Additional advantages of a UTS include what is likely a generalized upper bound for error [3], higher robustness, and high potential for low-SWaP computational cameras. Because of its low computational complexity, our approach in the future may be inverted to uncover the inverse mapping in data-driven models to solve inverse problems. A higher degree of sampling over the sensor images (i.e., zero-padding) may further reduce the reconstruction image error and even provide additional advantages, i.e., super-resolution phase retrieval from multiple phase-coded diffraction patterns [47] and depth detection [48].

Funding. Defense Advanced Research Projects Agency (YFA D19AP00036).

Acknowledgment. The authors acknowledge editing support from Ben Stewart (linkedin:benjamin-w-stewart).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," Nat. Photonics **12**, 578–589 (2018).
2. J. Wang, J. Liang, J. Cheng, Y. Guo, and L. Zeng, "Deep learning based image reconstruction algorithm for limited-angle translational computed tomography," PLOS ONE **15**, e0226963 (2020).
3. Y. Xue, S. Cheng, Y. Li, and L. Tian, "Reliable deep-learning-based phase imaging with uncertainty quantification," Optica **6**, 618–629 (2019).
4. R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A **7**, 394–411 (1990).
5. R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237–246 (1972).
6. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
7. Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," IEEE Signal Process. Mag. **32**, 87–109 (2015).
8. M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: thin, lensless cameras using coded aperture and computation," IEEE Trans. Comput. Imaging **3**, 384–397 (2016).
9. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," Optica **5**, 1–9 (2017).
10. K. Wakonig, A. Diaz, A. Bonnin, M. Stampanoni, A. Bergamaschi, J. Ihli, M. Guizar-Sicairos, and A. Menzel, "X-ray Fourier ptychography," Sci. Adv. **5**, eaav0282 (2019).
11. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**, 739–745 (2013).
12. P. C. Konda, L. Loetgering, K. C. Zhou, S. Xu, A. R. Harvey, and R. Horstmeyer, "Fourier ptychography: current applications and future promises," Opt. Express **28**, 9603–9630 (2020).
13. Y. Xiao, L. Zhou, and W. Chen, "Fourier spectrum retrieval in single-pixel imaging," IEEE Photonics J. **11**, 7800411 (2019).
14. M.-J. Sun and J.-M. Zhang, "Single-pixel imaging and its application in three-dimensional reconstruction: a brief review," Sensors **19**, 732 (2019).
15. M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," Nat. Photonics **13**, 13–20 (2018).
16. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," IEEE Signal Process. Mag. **25**, 83–91 (2008).
17. X. Hu, H. Zhang, Q. Zhao, P. Yu, Y. Li, and L. Gong, "Single-pixel phase imaging by Fourier spectrum sampling," Appl. Phys. Lett. **114**, 051102 (2019).
18. H. Deng, X. Gao, M. Ma, P. Yao, Q. Guan, X. Zhong, and J. Zhang, "Fourier single-pixel imaging using fewer illumination patterns," Appl. Phys. Lett. **114**, 221906 (2019).
19. Z. Liang, D. Yu, Z. Cheng, and X. Zhai, "Adaptive Fourier single-pixel imaging sampling based on frequency coefficients prediction," Opt. Eng. **59**, 073105 (2020).
20. Z. Zhang, X. Ma, and J. Zhong, "Single-pixel imaging by means of Fourier spectrum acquisition," Nat. Commun. **6**, 6225 (2015).
21. M. H. Seaberg, A. d'Aspremont, and J. J. Turner, "Coherent diffractive imaging using randomly coded masks," Appl. Phys. Lett. **107**, 231103 (2015).
22. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117–1125 (2017).
23. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181–1190 (2018).
24. C. A. Metzler, P. Schniter, A. Veeraraghavan, and R. G. Baraniuk, "prDeep: robust phase retrieval with a flexible deep network," in *Proceedings of the 35th International Conference on Machine Learning* (2018), pp. 3501–3510.
25. E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," Appl. Comput. Harmon. Anal. **39**, 277–299 (2015).
26. M. R. Kellman, E. Bostan, N. A. Repina, and L. Waller, "Physics-based learned design: optimized coded-illumination for quantitative phase imaging," IEEE Trans. Comput. Imaging **5**, 344–353 (2019).
27. C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," in *Proceedings of 5th International Conference on Learning Representations* (2016), pp. 1–15.
28. V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," Proc. Natl. Acad. Sci. USA **117**, 30088–30095 (2020).
29. B. Muminov and L. T. Vuong, "Fourier optical preprocessing in lieu of deep learning," Optica **7**, 1079–1088 (2020).
30. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," Opt. Express **24**, 13738–13743 (2016).
31. S. Malik, A. Sardana, and Jaya, "A keyless approach to image encryption," in *International Conference on Communication Systems and Network Technologies* (2012), pp. 879–883.
32. C. A. Metzler, F. Heide, P. Rangarajan, M. M. Balaji, A. Viswanath, A. Veeraraghavan, and R. G. Baraniuk, "Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging," Optica **7**, 63–71 (2020).
33. W. Luo, W. Alghamdi, and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," IEEE Trans. Signal Process. **67**, 2347–2356 (2019).
34. P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," IEEE Trans. Signal Process. **63**, 4814–4826 (2015).
35. Y. Luo, D. Mengu, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, and A. Ozcan, "Design of task-specific optical systems using broadband diffractive neural networks," Light Sci. Appl. **8**, 112 (2019).
36. E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, and Z. Yu, "Nanophotonic media for artificial neural inference," Photon. Res. **7**, 823–827 (2019).
37. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017).
38. G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," Nature **588**, 39–47 (2020).
39. D. Psaltis and N. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," Opt. Lett. **10**, 98–100 (1985).
40. S. Jutamulia and F. Yu, "Overview of hybrid optical neural networks," Opt. Laser Technol. **28**, 59–72 (1996).
41. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," Sci. Rep. **8**, 12324 (2018).

42. J. Wu, H. Zhang, W. Zhang, G. Jin, L. Cao, and G. Barbastathis, "Single-shot lensless imaging with Fresnel zone aperture and incoherent illumination," Light Sci. Appl. **9**, 53 (2020).
43. S. Zheng, X. Zeng, L. Zha, H. Shangguan, S. Xu, and D. Fan, "Orthogonality of diffractive deep neural networks," arXiv:1811.03370 (2018).
44. Q. R. Razlighi and N. Kehtarnavaz, "A comparison study of image spatial entropy," Proc. SPIE **7257**, 72571X (2009).
45. D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-8**, 413–424 (1986).
46. M. Deng, S. Li, Z. Zhang, I. Kang, N. X. Fang, and G. Barbastathis, "On the interplay between physical and content priors in deep learning for computational imaging," Opt. Express **28**, 24152–24170 (2020).
47. V. Katkovnik, I. Shevkunov, N. V. Petrov, and K. Egiazarian, "Computational super-resolution phase retrieval from multiple phase-coded diffraction patterns: simulation study and experiments," Optica **4**, 786–794 (2017).
48. Y. Hua, S. Nakamura, M. S. Asif, and A. C. Sankaranarayanan, "SweepCam—depth-aware lensless imaging using programmable masks," IEEE Trans. Pattern Anal. Mach. Intell. **42**, 1606–1617 (2020).