Research Article Vol. 7, No. 9 / September 2020 / Optica 1079

optica

Fourier optical preprocessing in lieu of deep learning

Baurzhan Muminov and Luat T. Vuong*

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

Received 15 May 2020; revised 10 July 2020; accepted 14 July 2020 (Doc. ID 397707); published 24 August 2020

Deep learning convolutional neural networks generally involve multiple-layer, forward-backward propagation machine-learning algorithms that are computationally costly. In this work, we demonstrate an alternative scheme to convolutional neural nets that reconstructs an original image from its optically preprocessed, Fourier-encoded pattern. The scheme is much less computationally demanding and more noise robust, and thus suited for high-speed and low-light imaging. We introduce a vortex phase transform with a lenslet-array to accompany shallow, dense, "small-brain" neural networks. Our single-shot coded-aperture approach exploits the coherent diffraction, compact representation, and edge enhancement of Fourier-transformed spiral phase gradients. With vortex encoding, a small brain is trained to deconvolve images at rates 5–20 times faster than those achieved with random encoding schemes, where greater advantages are gained in the presence of noise. Once trained, the small brain reconstructs an object from intensity-only data, solving an inverse mapping without performing iterations on each image and without deep learning schemes. With vortex Fourier encoding, we reconstruct MNIST Fashion objects illuminated with low-light flux (5 nJ/cm^2) at a rate of several thousand frames per second on a 15 W central processing unit. We demonstrate that Fourier optical preprocessing with vortex encoders achieves similar accuracies and speeds 2 orders of magnitude faster than convolutional neural networks. © 2020 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

https://doi.org/10.1364/OPTICA.397707

1. INTRODUCTION

Our ability to solve inverse problems and reconstruct object features from either incomplete or mixed-signal components is essential for a broad range of applications, from x ray imaging to remote sensing. Reconstruction or deconvolution of an object pattern from sensor data is often challenging from a practical standpoint, since algorithms must address the famous Phase Problem in which the phase information is lost by the sensor, which only registers photonic magnitude or intensity. Iterative approaches have been developed but are time-consuming, since the process may require multiple restarts with several initial guesses until convergence is achieved [1]. This area—iterative solutions of the Phase Problem—has developed and given rise to a set of optimization techniques that are today applied in many other domains [2]. Notably, these approaches provide the capacity to image through turbid and scattering media [3–5] and obtain depth estimation and all-focus images with coded apertures [6,7].

Recently, it has become possible to obviate the Phase Problem for image reconstruction with computational imaging. An expanding research area involves the application of neural networks, specifically deep learning convolutional neural networks (CNNs) [8]. With CNNs, the recording of an interference pattern such as a hologram, or several overlapping snapshots as with ptychography, can be used to reproduce object features [9–11]. When using the coherent diffraction through phase masks, at least two distinct images are generally needed to attempt the Phase Problem [12–14]. The first (to the authors' knowledge) application of CNNs for image reconstruction, is presented in [15], where a phase-encoded image on a spatial light modulator (SLM) is reconstructed via CNNs using intensity data from the camera. "Non-line-of-sight" CNN imaging was recently demonstrated from albedo autocorrelation patterns of speckles [16–18]. In these examples, as well as others that learn patterns without solving the Phase Problem, it is possible to reconstruct or predict an *object type* without being able to identify the *position* of the object [16].

Still, deep neural networks offer additional functionality in the process of reconstructing the object, for example, simultaneous autofocusing with phase recovery [19] or super-resolution in pixel-limited or diffraction-limited systems [20]. With sets of training and testing diffusers, the phase information encoded through controlled speckle patterns can be leveraged to predict the outputs from previously unseen diffusers [21]. The nonexhaustive list of important applications includes profilometry [22], imaging through smoke or strongly scattering media [5,23], and LIDAR that leverages multiple point cloud and time-of-flight information [24]. Additional examples of "nonlinear reservoir learning" are presented in [25], which employs caustic patterns for original object reconstruction. The challenges with deep learning methods include large training sets, long training times, and

slower execution speeds. These neural networks also have higher degrees of computational complexity that render them vulnerable to adversarial attacks [26].

In this paper, we focus on the application of shallow and dense neural networks and ask, is it possible to achieve additional image reconstruction functions without deep learning and without iterative schemes? Such "small-brain" approaches are regression-based and provide the advantage of a single forward pass, i.e., no requisite iterative phase retrieval procedures [27,28]. We demonstrate a new approach to image reconstruction with optical preprocessing in lieu of CNNs. Our strategy is similar to other hybrid and diffractive optical neural network approaches that aim to offload mathematical computation to the propagation of light [17,29–35]. What our scheme shows, unlike others, is that a simple neural network is capable of solving the inverse mapping with vortex spatial encoding in the Fourier domain. Moreover, the inverse mapping is performed efficiently and with less computational complexity with vortices than with random encoded patterns. This indicates that the optical vortex provides feature extraction in the Fourier representation, which further reduces the computational load. This article is an expanded version of work recently presented at [36].

Our approach provides new capacity to successfully image under low-light signal conditions. The results are dramatic since Fourier representations are compressed (i.e., the illuminated area of the camera sensor is much smaller than the area illuminated by the object's real image) [37] and robust (the resulting computer vision scheme is not susceptible to the rapid variations in scene illumination) [38]. We note that, from a purely computational standpoint, Fourier representations have been demonstrated to be efficient at solving classification problems [39–41]. Object reconstruction with Fourier representations reduces requisite memory, power, or energy requirements and may even achieve real-time image processing [42,43]. The advantages of Fourier operations further multiply since they may be completed optically before the digital neural network [32]. Still, in each of the aforementioned cases that use deep learning, the transferability of learned maps remains an issue—i.e., the trained neural nets are task-specific and, moreover, equipment-specific. This issue of transferability is further addressed in our work. We demonstrate that our "small-brain" approach does not require specificity in the trained data to solve the inverse problem. In fact, overcoming the Phase Problem with low computational complexity is our milestone result.

To achieve this, we exploit topological representations with optical vortices—more specifically, with Laguerre–Gaussian beams. Such beams with spiral phase gradients are characterized by a topological charge and associated with phase singularities at which the electric field is strictly zero [44–51]. A famous example that leverages phase singularities for imaging is the "vortex coronograph," in which a vortex phase is placed in the Fourier imaging plane. A higher-resolution vortex camera was recently demonstrated in [52], where the reconstruction contrast ratio is increased as a result of employing the vortex phase. The diffractive vortex optic is a phase encoder at the aperture, analogous to a grating for splitting and interfering multiple paths in single-shot Fourier ptychography schemes [53].

In our approach, we achieve noniterative, single-shot object reconstruction with a topological vortex-based lens array that contains multiple vortex phases in a lenslet pattern using the resulting edge-enhanced Fourier-plane representations. We explore the limits of the simplest neural nets and show that such neural nets are capable of generalizing the solution to a problem. We show that speed and robustness depend on the spatial phase encoding. The presence of the vortex provides spatial encoding to break the translation invariance of the measured Fourier pattern and solve the Phase Problem. We do not optimize the neural network for the highest quality of reconstruction, but rather to show the method's potential and its differences from other approaches. Image reconstruction is centered on dense neural nets or shallow neural nets. Again, we refer to this few-hidden-layer neural network that does not use deep learning as a "small-brain" [27].

2. INNOVATION

Figure 1 depicts our imaging scheme, where multiple images of the object $F(r, \phi)$ are collected in the Fourier domain: the light transmitted through each lenslet is modulated by different vortex and lens mask patterns $M_m(r, \phi)$; the camera detects the scaled, modulus-squared image of the Fresnel-propagated, vortex-Fourier-transformed intensity patterns, $|\tilde{F}_m(u, v)|^2$. Here, m is the vortex topological charge, r and ϕ are the real domain cylindrical coordinates, and u and v are the Fourier-plane Cartesian coordinates. The vortex Fourier intensity patterns $\tilde{F}$ are concentrated in a relatively small area but are typically donut-shaped with a wider donut as m increases [Fig. 1(b)]. The vortex phase in the object "real-domain" spatially encodes and breaks the translational invariance of the Fourier-transformed intensity pattern [Fig. 1(c)].

We consider a few small-image datasets as object inputs and compare different representations in $F(r, \phi)$. For each positive, real-valued dataset image $\mathbf{X}$, we map the phase changes,

$$F(r, \phi) = e^{i\alpha_0 \mathbf{X}}, \tag{1}$$

where α_0 is the dynamic range of the object phase-shift. This mapping is convenient because the signal power is invariant with our choice of $\mathbf{X}$. We have also considered opaque objects where $\mathbf{X}$

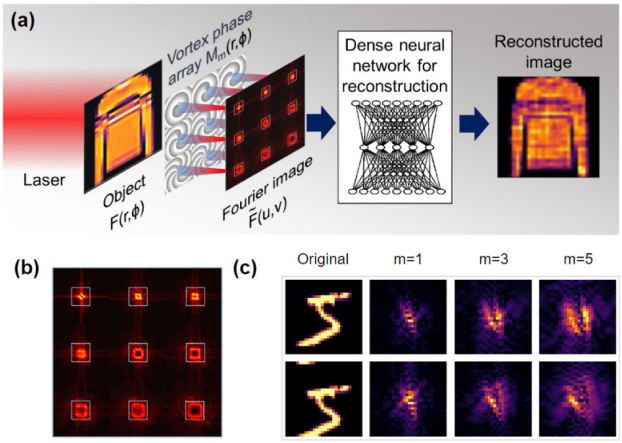

Fig. 1. (a) General schematic of the technique: a laser illuminates the object. Transmitted light is phase-modulated with a multivortex lens array. The back focal plane vortex Fourier intensity patterns are fed to a neural net that reconstructs the original image. (b) The vortex Fourier patterns have fewer pixels. Here, the combined area of nine dotted squares is equivalent to the area of the original object. (c) The vortex Fourier pattern for centered and shifted MNIST handwritten digit "5" show increasing sensitivity to shifts and larger areas (lower intensities) with increasing topological charges m.

blocks or absorbs the signal, i.e., $F(u, v) \propto \mathbf{X}$, which yields similar trends.

There are three primary innovations in our results. We demonstrate the following: (1) edge enhancement of spectral features with a vortex lens; (2) rapid inverse reconstruction of the image without a similar, learned dataset; and (3) robustness to noise, which depends on the layer activations.

A. Vortex Fourier Encoding and Feature Extraction

Here, we make two new claims about the special spatial qualities of optical vortices for image processing, namely:

- edge enhancement of the Fourier pattern, and
- compressed, coded fringes that mix the real and imaginary field components in the intensity measurements.

Consider a Fourier-space representation of a simple lenslet pattern composed of multiple-m vortex phases,

$$M_m(r, \phi) = \begin{cases} e^{\frac{-i\pi r^2}{\lambda f} + im\phi} & 0 \le r < a \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where a is the radius of the mask aperture, λ is the wavelength, and f is an effective focal length. This pattern with a centered vortex is appropriate for our dataset's mostly centered image objects $\mathbf{X}$. Our image reconstruction approach does not require that m is an integer. We show results of reconstructed images using fractional m in Fig. S1 in Supplement 1. At the phase plate [Eq. (2)], the transmitted pattern is a sum of Laguerre–Gaussian modes at $z = 0$ with different radial indices p,

$$M_m(r, \phi) = \sum_p W_p \mathrm{LG}_{m,p}(r, \phi), \tag{3}$$

where W_p represents modal coefficients related to index p and associated with Laguerre–Gaussian profiles, which we separate into components,

$$\mathrm{LG}_{m,p}(r, \phi) = L_p^{|m|}(2r^2/w^2) R(r) G(r) V_m(r, \phi), \tag{4}$$

$$R(r) = e^{\frac{-i\pi r^2}{\lambda f}}, \tag{5}$$

$$G(r) = \frac{1}{w} e^{-(\frac{r}{w})^2}, \tag{6}$$

$$V_m(r, \phi) = A_{m,p} r^{|m|} e^{im\phi}, \tag{7}$$

where $L_p^{|m|}(2r^2/w^2)$ are the generalized Laguerre polynomials that depend on m and p, and $A_{m,p} = \sqrt{\frac{2^{|m|+1} p!}{\pi (p+|m|)!}} w^{-|m|}$ [54]. We expect that the waist of the beam w is larger than the features of the object F. In our simulations, we fix w for all values of m, regardless of the training set so that the waist represents the aperture cutoff of a fixed Gaussian $G(r)$ beam regardless of m. In practice, if one consistently normalizes the mean-field radius of the transmitted vortex beam, then w would change with one's choice of both m and p.

The modal coefficients are

$$W_p = \int\int M_m \mathrm{LG}^*_{m,p}(r, \phi) r \mathrm{d}r \mathrm{d}\phi$$

$$= A_{m,p} \int_0^a 2\pi r^{|m|+1} L_p^{|m|}(2r^2/w^2) G(r) \mathrm{d}r. \tag{8}$$

The phase-singular term V_m is a radial magnitude gradient and azimuthal phase gradient, which can be simplified [55],

$$V_m(r, \phi) = A_{m,p} r^{|m|} e^{im\phi}$$

$$= A_{m,p} [r \cos(\phi) + ir \mathrm{sign}(m) \sin(\phi)]^{|m|}. \tag{9}$$

Since $r\cos(\phi)$ and $r\sin(\phi)$ are the canonical x and y Cartesian coordinates, which are Fourier transform pairs with $\frac{u}{f\lambda}$ and $\frac{v}{f\lambda}$ [56],

$$\tilde{V}_m(u, v) = \mathcal{F}\{V_m(r, \phi)\} = A_{m,p} \left(\lambda f \left[\mathrm{sign}(m) \frac{\partial}{\partial v} - i \frac{\partial}{\partial u} \right] \right)^{|m|}, \tag{10}$$

where $\mathcal{F}$ is the 2D Fourier transform operator. We view $\tilde{V}_m$ as a linear differential operator for the inputs to our neural network, which are the intensity patterns in the back focal plane [56],

$$\mathbf{Y} = |\tilde{F}_m(u, v)|^2 + \text{noise} \tag{11}$$

$$= |\sum_p \frac{W_p}{\lambda f} A_{m,p} \left(\lambda f \left[\mathrm{sign}(m) \frac{\partial}{\partial v} - i \frac{\partial}{\partial u} \right] \right)^{|m|}$$

$$\times \mathcal{F}\{F(r, \phi) L_p^{|m|}(2r^2/w^2) G(r)\}|^2 + \text{noise}. \tag{12}$$

In other words, the presence of the optical vortex provides unique preprocessing for the Fourier-plane data. The result is not purely from diffraction alone since the net diffraction from $V_m(r, \phi)$ cancels, or

$$\nabla_\perp^2 V_m(r, \phi) = A_{m,p} \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \phi^2} \right] r^{|m|} e^{im\phi} \tag{13}$$

$$= \left[\frac{|m|^2}{r^2} - \frac{m^2}{r^2} \right] V_m(r, \phi) \tag{14}$$

$$= 0. \tag{15}$$

Notably, the detected intensity patterns [Eq. (12)] are composed of real and imaginary differentials of the Fourier transform of F. The vortex-mixed real and imaginary field components produce asymmetric fringes in the intensity pattern. This asymmetry offered by the vortex phase is a chiral product of a radial change in magnitude multiplied by an azimuthal change in phase. This differential scheme provides feature extraction in a manner similar to that deployed in the HERALDO method for image reconstruction [57].

B. Small-Brain Inverse Reconstruction

Since a neural network is capable of guessing the reconstruction based on prelearned patterns without solving the inverse or Phase Problem, we take a new approach toward training and studying the

neural network. We test the inverse reconstruction of the neural network with several categorically patterned datasets, namely the Fashion-MNIST [58], Kuzushiji-MNIST [59], and Arabic [60], as well as the canonical handwritten MNIST digit dataset [61]. The "ground truth" outputs $\mathbf{X}$ are the dataset's 28×28 pixel images and are unit-normalized to provide comparable peak signal-to-noise (PSNR) with different image types across datasets. The vortex-based reconstruction achieves a mapping that is transferable or generalizable in cases when a random encoding scheme does not; if fast and accurate enough, the vortex imaging scheme may be capable of being applied as a "camera." To illustrate the range of potential, we show less accurate results when disjoint train-and-test datasets are used and more impressive image reconstruction when similar test-and-train datasets are used.

With our scheme, the inputs $\mathbf{Y}$ are the modulus-squared vortex Fourier transforms of Gaussian-apertured FG [Eq. (12)]. We set $\alpha_0 = \pi/2$ in a phase modulation scheme [Eq. (1)]. We set $f\lambda = 0.1$. If there is more than one vortex Fourier pattern used for reconstruction, the procedure is repeated, and the vortex images are catenated and/or truncated for the neural network input $\mathbf{Y}$. A dense, shallow, neural network with one hidden layer is trained with a mean-squared error (MSE) loss function. During training, the neural network is provided a subset of the related $\mathbf{X}$ and $\mathbf{Y}$, and during testing, the neural network is provided $\mathbf{Y}$ to solve for $\mathbf{X}$. The testing image set that the neural net has not seen before is referred to as the validation set.

The importance of spatial encoding in reconstruction is shown in Fig. 2. We train the model with numbers from the MNIST handwritten number dataset (flipped and inverted) [Fig. 2(a)]. However, we test with a separate dataset with patterns that are different. With Kuzushiji characters [Fig. 2(b)], the reconstruction fails when there is no spatial encoding and also fails with a random spatial encoder [Figs. 2(c) and 2(d)], but succeeds, with some loss of resolution, with a vortex Fourier encoder ($m = 3$) [Fig. 2(e)]. With the Arabic data set [Figs. 2(f) and 2(g)], the reconstructed letters are impressive considering that we limit our training to the types of handwritten digits that deviate substantially from the formal Arabic letters. This illustration shows one approach to testing our intuition about the Phase Problem with neural networks and also demonstrates the unique opportunities with vortex Fourier encoding schemes: a combination of compressed, encoded inputs is critical. Without both compressed and encoded inputs and without previous patterns for guessing, the neural network cannot produce an inverse map from $\mathbf{Y}$ to $\mathbf{X}$.

The reconstruction quality of the images in Fig. 2 is not impressive because we have chosen an arbitrary training set that is unrelated to the test set. With a better training set, it should be possible to solve the inverse problem generally. We repeat the machine-learning problem with the Fashion-MNIST dataset, where the training and testing sets are more similar, and where, subsequently, the neural network is able to provide the inverse mapping of the $\mathbf{Y}$ to $\mathbf{X}$ without spatial encoding. Figure 3(a) shows the validation set. Even though the neural network has not seen the validation set before, unlike in the previous example, it has been trained with similar sets of images that fall into various categories (shirts, shoes, dresses, etc.). Figure 3(b) shows discernible reconstructions of images without spatial encoding. In this case, the neural network has learned and reconstructed patterns. The reconstructed images exhibit ghosting as a result of this uncertainty. Again, it is important to emphasize that even though the neural network is able to reconstruct the MNIST images, it does so with learned similarities with the training set, which is not an inverse mapping.

Fig. 2. (a) Thirty-six images of the training set composed of MNIST handwritten digits flipped vertically and horizontally. (b) Twenty-five images of the MNIST Kuzushiji test set. Reconstructed images with (c) no spatial encoding, (d) random spatial encoding, and (e) vortex spatial encoding with $m = 3$. When we train with the same MNIST digits in (a) and (f) test with Arabic letters, (g) the reconstructed images with vortex Fourier encoding and $m = 3$ are in good agreement. With vortex Fourier encoding, the neural network produces a generalizable inverse map for image reconstruction.

With a vortex pattern, the Fourier-image phase is preserved and encoded, and the reconstructed images are impressively sharp and delineated. This is an important remark as the ghosting or faded silhouettes are problematic for classification and computer vision algorithms [62]. In addition to MSE, we employ the structural similarity index metrics (SSIM) to quantify the reconstruction quality. The SSIM is limited to a [0,1] segment, which is a more reasonable metric for human quality perception evaluation [63]. This difference between SSIM and MSE illustrated Fig. S2 Supplement 1.

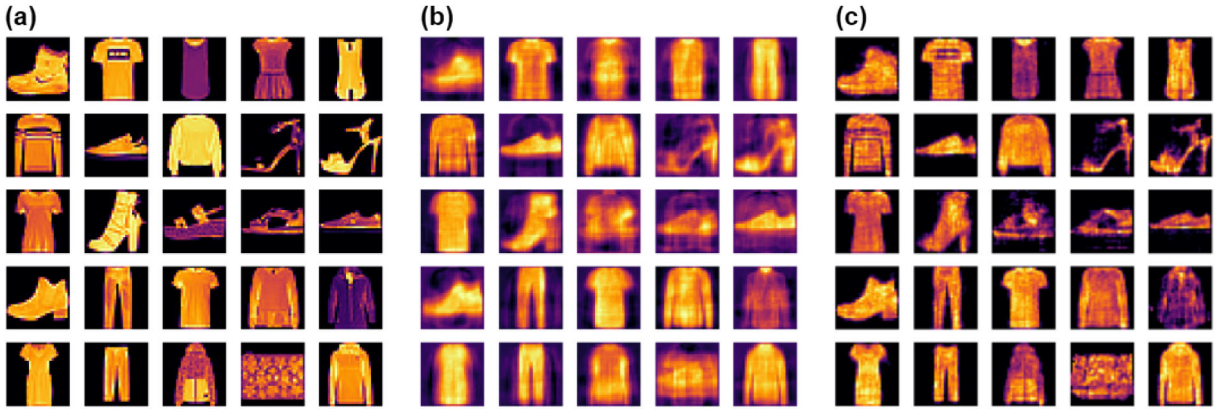

Fig. 3. (a) Twenty-five of the original Fashion-MNIST dataset images. Reconstruction with (b) $m = 0$, with no spatial encoding and (c) two vortices of topological charges $m = 1, 3$ with nonlinear activation used in the last layer. Without spatial encoding, the neural network still learns the patterns when there are categorical variations between training and test sets. The structural similarity index metric (SSIM) is quantified in Table 1.

Table 1. SSIM and MSE for Fashion-MNIST Reconstruction with a Small-Brain (SB) and Convolutional Neural Networks (CNN)[a]

	SB, No Vortex	SB, 1 Vortex	SB, 2 Vortices	SB, 3 Vortices	CNN, 1 Vortex	CNN, 2 Vortices (U-net)
SSIM	0.45	0.62	0.84	0.88	0.61	0.84 (0.87)
MSE	0.0280	0.0242	0.0140	0.0122	0.0235	0.0145 (0.0124)
Testing speed	8600 FPS	8600 FPS	6400 FPS	5600 FPS	200 FPS	92 FPS (27 FPS)
Training time	2 min	2 min	5 min	7 min	35 min	78 min (440 min)

[a]U-net CNN architecture is also implemented; near-optimized quality is achieved with two vortices.

Table 1 illustrates the convergence of the reconstructed images of the Fashion-MNIST dataset to the original with SSIM, MSE, and speed for different numbers of vortices. For comparison, Table 1 also shows the SSIM, MSE, and speeds for a three-layer CNN-trained reconstruction with single and dual-vortex datasets (the result for seven layer U-net is shown in parenthesis [64]). The reconstruction speed is measured using precollected data, and we only consider the time for the reconstruction: the CPU time needed for Fourier transform is not included in the speed calculations, since in practice we expect that the vortex Fourier encoding is completed optically. Table 1 suggests that our proposed architecture achieves the same quality with much lower computational overhead; on the same CPU, our approach reconstructs at a rate of more than 6000 images per second (FPS) while a three-layer CNN with 3×3 kernel achieves only 90 FPS. We also provide training times for different models, which highlight at least an order-of-magnitude advantage of the small brain over the simplest CNNs.

If we use the detected patterns from four and five vortices as inputs to the neural network, we encounter increasing accuracy but diminishing returns. For example, three vortices achieve SSIM $= 0.88$, four vortices achieve SSIM $= 0.89$, and five vortices achieve SSIM $= 0.91$. The results for comparison using a simple CNN with a 3×3 kernel achieve similar SSIM with the same inputs. Additionally, we implement a U-net multiple-layer CNN [64] with two vortices. While additional layers yield increased SSIM, this improvement comes with increased computational complexity and memory requirements. We limit the scope of this paper to simpler neural networks for speed and robustness.

C. Speed and Robustness to Noise

We study the variable speed and robustness of the reconstruction with vortex Fourier encoding with other random encoding approaches and consider sensor shot and dark noise in the neural network input [Eq. (12)] [65],

$$\text{noise} = P_n(|\tilde{F}(u, v)|^2) + P_n(\sigma_d^2). \tag{16}$$

Both noise terms have Poisson distributions $P_n(\mu)$, where μ is the expected value and variance. The variance of the sensor noise is proportional to the intensity over the pixel $|\tilde{F}(u, v)|^2$ while the variance of the dark noise σ_d^2 is related to the dark current and read noise of each pixel. The continuously valued noise is related to the camera noise with MSE,

$$\text{MSE} = \frac{1}{N}\Sigma(y_0 - y_i)^2, \tag{17}$$

where N is the number of pixels, and y_0 and y_i are the noiseless and noisy pixels of $\mathbf{Y}$.

For our simulations here, we assume that the specifications of the camera do not interfere with our reconstruction algorithm. Therefore, regardless of the light intensity, we assume that the full dynamic range of a 12-bit camera is used. To vary the noise, we keep σ_d^2 fixed and change $|\tilde{F}(u, v)|^2$ to study the PSNR [63],

$$\text{PSNR} = 10\log_{10}\frac{P_{\text{pk,signal}}}{P_{\text{noise}}} \tag{18}$$

$$= 10\log_{10}\left(\frac{(2^L - 1)^2}{\text{MSE}}\right) \tag{19}$$

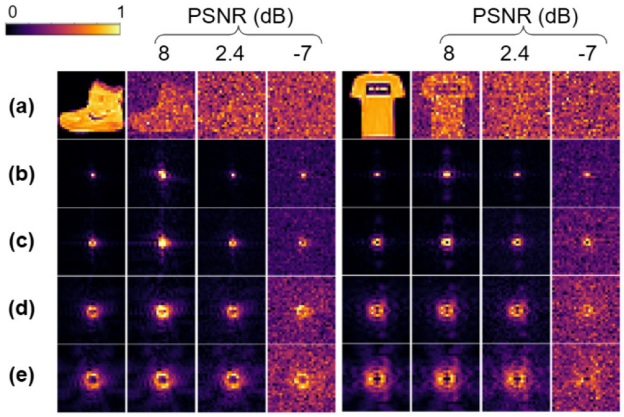

Fig. 4. Normalized test inputs **Y** to the neural network with noise for two Fashion-MNIST images. Each block shows decreasing peak signal-to-noise ratio (PSNR) in columns from left to right. (a) Object data **X** and vortex spatial encoding (b) $m = 0$ or no vortex (c) $m = 1$, (d) $m = 3$, and (e) $m = 5$. The Fourier transform representations have higher PSNR that decreases with higher m given the same camera and light flux for **X**.

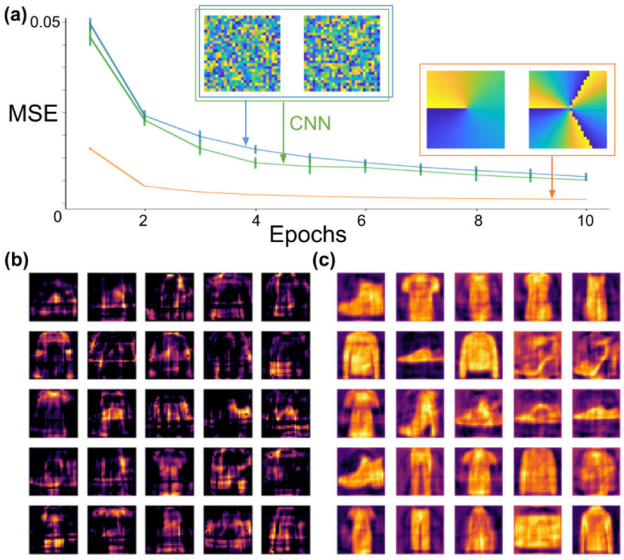

Fig. 5. (a) Comparison of training rates (MSE versus epochs) for different encoders in the reconstruction of MNIST fashion images. Dual-vortex encoding $m = 1, 3$ (red, small-brain) converges faster than random-pattern encoding, which has a similar rate of convergence for a small-brain (blue) and convolutional neural network (green). The reconstructed images with (b) random-pattern and (c) vortex encoders with 2 dB PSNR.

$$= 20\log_{10} \frac{\max(|\tilde{F}(u, v)|^2)}{\langle |\tilde{F}(u, v)|^2 \rangle + \sigma_d^2}, \tag{20}$$

where $P_{\text{pk,signal}}$ is the peak power over the camera detector, P_{noise} is the average power in the noise, and the camera dynamic range is denoted by $L(L = 12$ for a 12-bit camera). The additional factor of 2 in Eq. (20) arises because the electrical PSNR is proportional to the camera detector voltage, squared, while the voltage is proportional to the electric field, squared. We continuously vary the electrical pixel power in images to change the PSNR. Meanwhile, as mentioned above, we discretize the inputs to the neural network to use the maximum range of the 12-bit camera.

Figure 4 illustrates the trade-off between resolution and robustness to noise. With $m = 0$ when the majority of the power is on-axis and there is no vortex, the signal intensity is most robust to noise, but the neural network cannot solve the Phase Problem. With higher m (as well as larger $f\lambda$), the Fourier-plane pattern covers a larger area, so that the spectral features are sampled with better resolution. At the same time, when the area is larger, the effective PSNR decreases, resulting in a Fourier representation less robust to noise.

Both with and without noise, we compare our results to a random spatial encoding pattern, where vortices are replaced with a diffuser, for example [25], and imaged in the Fourier plane. As with other spatial encoding schemes, the SSIM in reconstruction using the random phase patterns approaches the level of performance of vortex schemes in the no-noise scene when similar images are used to train and test the dataset. However, in order to achieve "near-vortex" performance, the random encoder requires more training time, which increases from 3 epochs to 8–10 epochs without noise. Furthermore, the situation changes drastically as we enter the noisy regime: while the vortex Fourier encoding preserves image quality in the case of high noise, the random encoding scheme fails completely. The gains with accuracy also increase with the vortex encoding scheme, as shown in Fig. 5 with 2 dB PSNR. Thus, vortex encoding provides feature extraction for efficient reconstruction of the object **Y** to **X**, which enables faster convergence of the neural network, as well as robust construction in the presence of noise. An interesting aspect of the inverse problem arises in the presence of noise when the neural network is trained with images without noise and is also tested with images from the same dataset with noise. (Again, only two vortices are needed to produce the near-ideal reconstructed images.) The neural network learns an inverse mapping that minimizes MSE in the training set; however, different layer activations are vulnerable to different types of noise. This is illustrated in Fig. 6. A linear activation leads to more ghosting and amplification of sensor shot noise, whereas a nonlinear activation is more vulnerable to the dark noise. While the nonlinear activation produces sharper images in the absence of noise, the linear activation is more robust and produces better images in the presence of more noise. With high noise, the nonlinear model's image quality is mixed. We find that the most robust model uses linear activations in the hidden layer and nonlinear activations only for the last layer. The linear activations produce results that are more generalizable and transferable in the presence of noise. In contrast, while more accurate with low noise, nonlinear activations provide less of the "inverse" mapping, as seen by a bias in reconstruction for highlighting edges. The CNN performs better in the case of lower noise levels, but cannot produce any meaningful features in the high noise regime. The all-linear activation model suffers from significant Gaussian blur or ghosting compared to the model with linear-nonlinear activation. This suggests that the nonlinear layer filters the Gaussian blur, which is reliably used at the last stage.

A comparison of different random patterns as spatial encoders suggests that the use of completely stochastic patterns are not most suited for image reconstruction and the structured base sets, such as topological phases and vortices, are more effective for this task. This said, the work in [25] is often applied to incoherent light illumination for image processing. There is certainly space for further study of topological structures for light scattering that connects different spatial-encoding and compressive imaging approaches with neural networks.

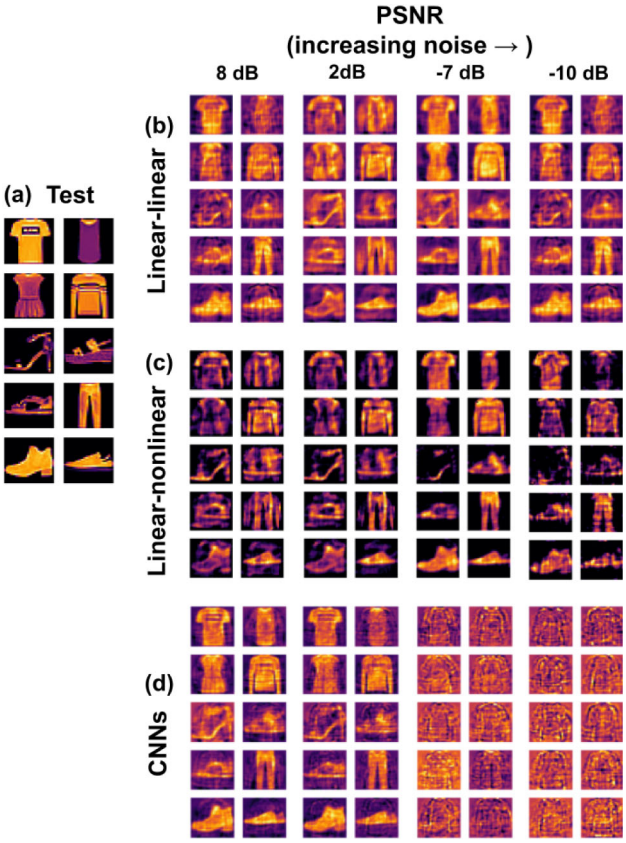

Fig. 6. (a) Validation test images. Reconstruction results under PSNR-labeled noisy conditions that show the effect of (b) "linear" and (c) "nonlinear" activations for the final layer of the small-brain or (d) convolutional neural network. A vortex encoder with $m = 1, 3$ is used.

3. EXPERIMENTS AND ANALYSIS

In experiments, we capture Fourier-plane intensity patterns when the object illumination flux is too low to be imaged directly by the camera. In this way, the vortex imaging and small-brain deconvolution approach serves as a low-light-level camera, suitable for imaging through noisy environments.

The experimental setup is shown in Fig. 7(a). Coherent polarized light from the laser source (diameter, 1.1 mm) is reflected from the two mirrors, passes through the half-wave plate and a second polarizer, which rotates to vary the power. The laser is a 500 fs pulsed Nd:YAG Fianium at its second harmonic, $\lambda = 532$ nm. A spatial filter eliminates parasitic modes. The transmitted, linearly polarized $TEM_{0,0}$ beam is collimated (diameter, 1.5 cm) onto the SLM (Hammamatsu LCOS-SLM) at an incident angle of 30°. The SLM has 800×600 pixels. Original images were upscaled in order to fill the SLM matrix size, so each spiral phase pattern is $\tilde{1}65 \times 165$ pixels. The neural network used for two vortex inputs is shown in Fig. 7(b).

When recording experimental data, the phase patterns are saved in files at the start. An example phase pattern is shown in Fig. 7(c). Each pattern contains six lenslet images with vorticial ($m = 4 - 9$) and quadratic radial phases ($f/\lambda = 0.1$), as well as the imprint of an MNIST handwritten digit. The quadratic phase is an artificial lens for focusing and replaces a lens array, so we can collect all six images simultaneously. An automated computer program sends the phase pattern from the saved file to the SLM, while a second program records the reflected light and grabs the CCD camera image. The CCD camera (Thorlabs-DCU223M) is approximately 20 cm from the SLM and has 4.65 μm × 4.65-μm pixel area, 8-bit dynamic range, and 1024×768 pixel resolution. The oblique reflections of the SLM phase pattern result in vertically elongated images on the CCD.

We achieve impressive results with this 8-bit CCD camera. Imaging is demonstrated here with average light fluxes of 30 nJ/cm^2 (intensities of 10 μW/cm^2, exposure times of 2.8 ms) over six vortex masks at the SLM. A "reference" CCD image that is taken with higher intensity is shown in Fig. 7(d). This Fourier-plane image is used to center the cropped data since with such low illumination fluxes, we use only 5–10% of the dynamic range of the 8-bit camera, and the images are virtually all dark. For comparison, the signal flux at the SLM would be below the level

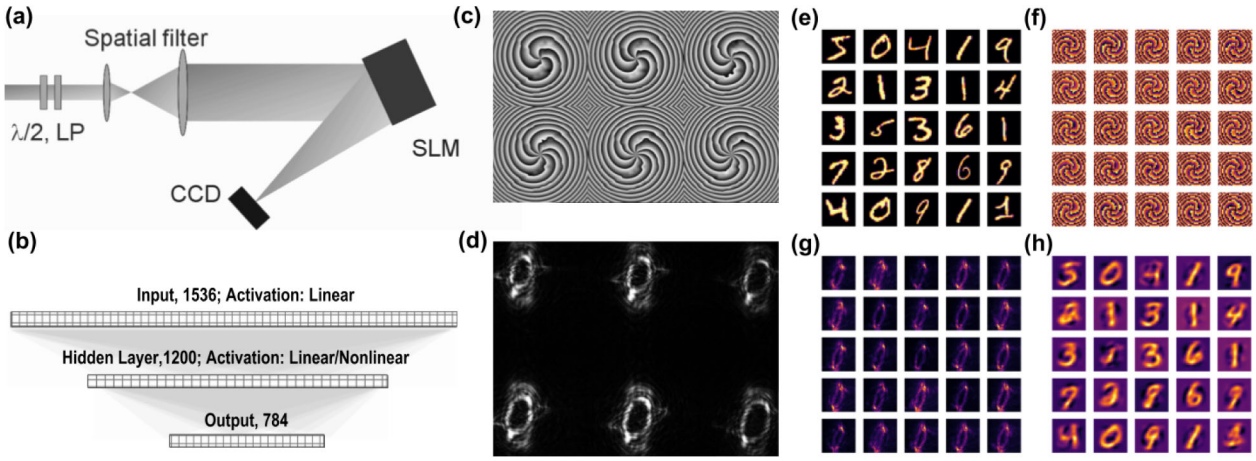

Fig. 7. (a) Experimental setup. A laser ($\lambda = 532$ nm) is directed through the half-wave plate ($\lambda/2$), linear polarizer (LP), and spatial filter. The beam is collimated and illuminates the reflective spatial light modulator (SLM). A PC supplies patterns to the SLM. These vortex-object patterns also have a quadratic phase to focus all six patterns on the CCD camera, whose images are stored on the same PC. (b) The neural net architecture used for reconstruction. (c) Sample SLM pattern that imprints $m = 4 - 9$ vortices, a quadratic radial phase, and MNIST handwritten digit from the PC. (d) Example CCD image for vortices $m = 4 - 9$. (e) Twenty-five test images, (f) the corresponding 25 phase patterns with $m = 4$, (g) resulting CCD images with $m = 4$, and (h) reconstructed images.

of the camera read noise, even with minute-long camera exposure intervals.

With our automated MATLAB program, we record approximately five images per second for the training and testing of our algorithm. The primary limitation with speed is the camera retrieval time, which is much longer than the camera shutter time. The delay time between image uploads and the presence of SLM phase patterns is 50 ms. To achieve experimental results analogous to the simulated efforts, the CCD image subsets are cropped and downsized with "inter-area" interpolation to 28×28 pixels for each $\mathbf{Y}$. A square area is cropped. The validation/test, reconstructed, and low-flux SLM/CCD images (10 μW power and 2.8 ms exposure) are shown in Figs. 7(e) and 7(h). All images are normalized to the color-range since without this color-bar normalization, the vortex Fourier patterns would not be visible. Even though there are six vortex lenslets in the CCD images, we use only one vortex $m = 4$ and achieve an SSIM of 0.688 with all-linear and linear-nonlinear activations. The choice of m did not significantly change results. We use only 4500 (500) images from the dataset for training (testing) even though there are 60,000 images in the MNIST dataset. In fact, the training of the algorithm converges to similar values of accuracy, without overfitting, with only 2000 training images.

While we could integrate more images from the dataset or more patterns from different-m vortices to train the small brain, doing so would not improve the accuracy of the reconstruction algorithm, which at this low-light level is limited by the camera characteristics. Similarly, we do not see significant improvement by varying the activations of the neural network layers. In other words, the primary limitation is the sensitivity of our camera. In this low-light level range, we are using only 4 bits of the camera (unlike in simulations where we assumed the full dynamic range of a 12-bit camera). While the robustness to overfitting in the training algorithm is impressive, i.e., we need only 2000 images to train the algorithm, the training time increases by a factor of 4 compared to simulations. Longer training times may arise when the MSE gradients are shallow and when the dynamic range of the neural network inputs are limited.

Once trained, we still reconstruct MNIST Fashion images from a vortex Fourier representation at a rate of several thousand frames per second on a 15 W central processing unit, 2 orders of magnitude faster than CNNs. Our approach to and understanding of image reconstruction with the topological Fourier encoding and small-brain neural network is new. Future work should include the effects of different types of noise, variations between training and test sets, topological phase and compression, and camera characteristics on the robustness of the neural network reconstruction.

We emphasize that the main advantages of our "small-brain" approach are high speeds, robustness to noise, and the demonstrated ability for generalization—which opens an avenue for a camera scheme, where the vortex phase is at the lens of a camera [52]. In a camera scheme, a real object and fixed vortex encoding aperture would replace the SLM. In such an imaging system, the bottleneck would the camera or processing speed rather than the SLM, which is a replacement in preliminary studies.

In actuality, a real system with a vortex-coded aperture may also exhibit misalignment. We simulate the misalignment effect by randomly shifting the raw test images by several pixels, while leaving the training set images intact. The neural net built on raw experimental data (75×75 pixels) seems to exploit the fine structure of the Fourier pattern; on the other hand, the model that uses the downsized images (28×28 pixels) relies on the continuous patterns. While the neural network trained on the raw, unshifted data suffers significantly—SSIM drops by a factor of 4 even after single pixel shift—the NN built using a downsized image is much more robust: for [1, 2, 3, 45] pixel jitter during testing, the SSIM degrades [4, 7, 20, 39, 78]%. This result may indicate some value of downsizing data for robustness to alignment issues.

Finally, there are some differences between simulations and experiments worth mentioning, particularly since artifacts may favorably enhance the results gained from experiments. We use an SLM with higher resolution than the 28×28 images used in experiments, and additionally, the camera images have higher resolution when cropped and are subsequently downsized. The higher-resolution SLM and camera images may have two consequences: the vortex is imprinted with higher resolution than the object in simulations; additionally, there is likely scattering from the SLM pixels, which interfere and increase the information in the downsized images. The downsizing algorithm does not affect the performance as long as the neural network is retrained with the same downsizing approach; however, it may retain information of the higher-pixel SLM and camera resolution. We consistently achieve higher reconstruction quality with the SLM than with numerical simulations. This experimental observation indicates that there is information to be retrieved through scattering with hybrid machine vision systems.

4. SUMMARY AND CONCLUSION

We present a vortex Fourier encoding approach to preprocessing data prior to a neural network. The phase singularity is imprinted on the object prior to the lens Fourier transform. As a result of the spiral phase, we observe edge-enhanced, compressed, and phase-preserved representations. While many inverse problems are solved iteratively or involve CNNs, we obviate the need for CNNs and show that it is possible to solve inverse problems with Fourier representations of vortex-encoded objects. Spatial encoding with a topological phase results in efficient feature extraction of the Fourier pattern and accelerates learning for the inverse reconstruction of the object. Our approach, using shallow, dense neural networks or "small-brain" machine learning, offers a strategy for accurate, robust, and rapid camera-like imaging in low-light or noisy environments.

The reconstruction quality from the vortex-phase coded-aperture patterns is high when the test and training sets are similar for on-axis centered objects. When objects are located off-axis or when the mask phase singularity does not overlap with the light that is transmitted through the object, the reconstruction quality decreases. Reconstruction quality is related to the distinct interference fringe patterns that arise when light from the object passes through opposite sides of the phase singularity. Since vortex phase patterns represent a superposition of Laguerre–Gaussian modes, the task of analyzing multiple, superposed, topological patterns with off-axis vortices is relatively straightforward. Phase masks with additional off-axis vortices would capture greater information from off-axis objects.

The specific spatial encoding by a vortex provides mixed edge-enhanced real and imaginary components. With vortex encoding, we find a sparse neural network is capable of being transferable and generalizable. We aim to unbox aspects of the black box of machine

learning by solving the generalized problem of inverse reconstruction with categorically similar and dissimilar images and different MNIST datasets. We also show that there are different levels of robustness to different types of noise/deviation from training sets with different layer activations. Nevertheless, our small-brain machine-learning algorithm reduces the computational overhead with training and also reduces computational complexity in reconstruction. Given that more complicated neural network architectures can be unpredictable, [26], our approach that provides the one-to-one map is appropriate for mission-critical problems, security-related systems, or autonomous vehicles.

The reconstruction quality from the vortex-phase coded-aperture patterns is high when the test and training sets are similar for on-axis centered objects. When objects are located off-axis or when the mask phase singularity does not overlap with the light that is transmitted through the object, the reconstruction quality decreases. Reconstruction quality is related to the distinct interference fringe patterns that arise when light from the object passes through opposite sides of the phase singularity. Since vortex phase patterns represent a superposition of Laguerre–Gaussian modes as shown above, the task of analyzing multiple, superposed, topological patterns with off-axis vortices is relatively straightforward. Phase masks with additional off-axis vortices would capture greater information from off-axis objects.

To summarize, the optical preprocessing approach demonstrated here with a topological phase mask and lens is:

- **Robust to noise.** Signals effectively achieve 200–2000X higher PSNRs. We successfully capture and deconvolve objects illuminated with 10 $\mu W/cm^2$ average intensities with millisecond shutter times using 4 bits of an 8-bit CCD camera.
- **Single shot.** Reconstruction is possible with a single image containing two vortices or two orthogonal topological phases.
- **Low-latency and fast.** Our approach has potential for real-time processing and video-camera streaming. With Fashion-MNIST images, we process several thousand frames per second with low-power hardware (10–20 W).
- **Computationally efficient.** While other methods currently take multiple encoded images or use iterative schemes, we achieve near-to-ideal reconstruction with two phase-encoded images.
- **Extremely low-power computation.** The technique uses explicitly simple neural nets (no deep learning) where preprocessing is completed with parallel optical propagation.
- **Compact with memory.** The vortex Fourier transform provides a compressed representation that can be leveraged to minimize the number of pixels that carry data forward.
- **Flexible with a digital readjustable stage.** There is a trade-off between resolution, robustness, and sampling that we control with the choice of vortex charge m and focal length f.

If this approach is successful at reconstructing real objects with depth of field, there are numerous applications that involve imaging in low signal conditions, such as security-related systems where illumination is minimal, driver-assist systems, and microscopy of delicate photosensitive biological samples, among others. Given the low power requirements and high frame-rate reconstruction speeds, our scheme is expected to be useful for satellite or unmanned operations. The vortex Fourier encoding scheme may be efficient at collecting radiation in pulsed, spectroscopic, laser applications. While this paper was in review, a similar demonstration of "small-brain" image reconstruction from light in multimode fiber was reported [66], and it is interesting to see how the simple feed-forward network outperforms the U-net CNNs with spectral (speckle) image inputs. Since many higher-order fiber modes carry phase singularities, multimode fibers may provide another means for vortex encoding. Further research with free-space imaging may exploit topological features for achieving greater depth of field [67,68] and extend our knowledge of vortex Fourier encoders to leverage information in light polarization, dispersion, and spatiotemporal coherence.

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.

See Supplement 1 for supporting content.

REFERENCES

1. T. Latychevskaia, "Iterative phase retrieval in coherent diffractive imaging: practical issues," Appl. Opt. **57**, 7187 (2018).
2. V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," Proc. Natl. Acad. Sci. USA **104**, 418–423 (2007).
3. J. Bertolotti, E. G. van Putten, C. Blum, A. Lagendijk, W. L. Vos, and A. P. Mosk, "Non-invasive imaging through opaque scattering layers," Nature **491**, 232–234 (2012).
4. O. Katz, E. Small, Y. Guan, and Y. Silberberg, "Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers," Optica **1**, 170 (2014).
5. A. Turpin, I. Vishniakou, and J. D. Seelig, "Light scattering control in transmission and reflection with neural networks," Opt. Express **26**, 30911 (2018).
6. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," ACM Trans. Graph **26**, 70 (2007).
7. W. Chi and N. George, "Phase-coded aperture for optical imaging," Opt. Commun. **282**, 2110–2117 (2009).
8. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**, 921 (2019).
9. T. Tahara, X. Quan, R. Otani, Y. Takaki, and O. Matoba, "Digital holography and its multidimensional imaging applications: a review," Microscopy **67**, 55–67 (2018).
10. Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," Light Sci. Appl. **8**, 1–8 (2019).
11. 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).
12. P. Sidorenko and O. Cohen, "Single-shot ptychography," Optica **3**, 9–14 (2016).
13. F. Zhang, I. Peterson, J. Vila-Comamala, A. Diaz, F. Berenguer, R. Bean, B. Chen, A. Menzel, I. K. Robinson, and J. M. Rodenburg, "Translation position determination in ptychographic coherent diffraction imaging," Opt. Express **21**, 13592–13606 (2013).
14. H. Wang and R. Piestun, "Azimuthal multiplexing 3D diffractive optics," Sci. Rep. **10**, 6438 (2020).
15. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117–1125 (2017).
16. 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).
17. M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**, 338–341 (2018).

18. X. Lei, L. He, Y. Tan, K. X. Wang, X. Wang, Y. Du, S. Fan, and Z. Yu, "Direct object recognition without line-of-sight using optical coherence," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2019), pp. 11737–11746.
19. Y. Wu, Y. Rivenson, Y. Zhang, Z. Wei, H. Günaydin, X. Lin, and A. Ozcan, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," Optica **5**, 704–710 (2018).
20. T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," Sci. Rep. **9**, 3926 (2019).
21. 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).
22. K. Song, S. Hu, X. Wen, and Y. Yan, "Fast 3D shape measurement using Fourier transform profilometry without phase unwrapping," Opt. Lasers Eng. **84**, 74–81 (2016).
23. M. Locatelli, E. Pugliese, M. Paturzo, V. Bianco, A. Finizio, A. Pelagotti, P. Poggi, L. Miccio, R. Meucci, and P. Ferraro, "Imaging live humans through smoke and flames using far-infrared digital holography," Opt. Express **21**, 5379–5390 (2013).
24. E. W. Mitchell, M. S. Hoehler, F. R. Giorgetta, T. Hayden, G. B. Rieker, N. R. Newbury, and E. Baumann, "Coherent laser ranging for precision imaging through flames," Optica **5**, 988–995 (2018).
25. 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 (2018).
26. I. Goodfellow, P. McDaniel, and N. Papernot, "Making machine learning robust against adversarial inputs," Commun. ACM **61**, 56–66 (2018).
27. L. Vuong and H. Lane, "Nonlinear spectral preprocessing for small-brain machine learning," Proc. SPIE **11139**, 111390T (2019).
28. R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," Opt. Express **24**, 13738–13743 (2016).
29. Y. Zhou, H. Zheng, I. I. Kravchenko, and J. Valentine, "Flat optics for image differentiation," Nat. Photonics **14**, 316–323 (2020).
30. 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).
31. A. Cordaro, H. Kwon, D. Sounas, A. F. Koenderink, A. Alù, and A. Polman, "High-index dielectric metasurfaces performing mathematical operations," Nano Lett. **19**, 8418–8423 (2019).
32. 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).
33. T. Zhu, Y. Zhou, Y. Lou, H. Ye, M. Qiu, Z. Ruan, and S. Fan, "Plasmonic computing of spatial differentiation," Nat. Commun. **8**, 12324 (2017).
34. C. Guo, M. Xiao, M. Minkov, Y. Shi, and S. Fan, "Photonic crystal slab Laplace operator for image differentiation," Optica **5**, 251–256 (2018).
35. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–1008 (2018).
36. B. Muminov and L. T. Vuong, "Fourier-plane vortex laser holography for robust, small-brain machine learning and image classification," in *Conference on Lasers and Electro-Optics* (OSA, 2020).
37. R. M. Balboa and N. M. Grzywacz, "Power spectra and distribution of contrasts of natural images from different habitats," Vis. Res. **43**, 2527–2537 (2003).
38. F. Krahmer and R. Ward, "Stable and robust sampling strategies for compressive imaging," IEEE Trans. Image Process. **23**, 612–622 (2014).
39. O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *29th Annual Conference on Neural Information Processing Systems (NIPS), Advances in Neural Information Processing Systems* , C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, eds., (Vol. **28**), (2015).
40. M. Alberti, A. Botros, N. Schuez, R. Ingold, M. Liwicki, and M. Seuret, "Trainable spectrally initializable matrix transformations in convolutional neural networks," arXiv:1911.05045 (2019).
41. C.-A. Popa and C. Cernăzanu-Glăvan, "Fourier transform-based image classification using complex-valued convolutional neural networks," in *Advances in Neural Networks—ISNN* (Springer, 2018), pp. 300–309.
42. W. Chen, J. Wilson, S. Tyree, K. Q. Weinberger, and Y. Chen, "Compressing convolutional neural networks in the frequency domain," in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD'16* (ACM, 2016).
43. M. J. Cherukara, Y. S. G. Nashed, and R. J. Harder, "Real-time coherent diffraction inversion using deep generative networks," Sci. Rep. **8**, 16520 (2018).
44. X. Wang, Z. Nie, Y. Liang, J. Wang, T. Li, and B. Jia, "Recent advances on optical vortex generation," Nanophotonics **7**, 1533–1556 (2018).
45. L. Allen, M. W. Beijersbergen, R. J. C. Spreeuw, and J. P. Woerdman, "Orbital angular momentum of light and the transformation of Laguerre–Gaussian laser modes," Phys. Rev. A **45**, 8185–8189 (1992).
46. M. S. Soskin, V. N. Gorshkov, M. V. Vasnetsov, J. T. Malos, and N. R. Heckenberg, "Topological charge and angular momentum of light beams carrying optical vortices," Phys. Rev. A **56**, 4064–4075 (1997).
47. J. H. Lee, G. Foo, E. G. Johnson, and G. A. Swartzlander, "Experimental verification of an optical vortex coronagraph," Phys. Rev. Lett. **97**, 053901 (2006).
48. M. R. Dennis, K. O'Holleran, and M. J. Padgett, "Chapter 5 singular optics: optical vortices and polarization singularities," in *Progress in Optics* (Elsevier, 2009), pp. 293–363.
49. L. Zhu and J. Wang, "A review of multiple optical vortices generation: methods and applications," Front. Optoelectron. **12**, 52–68 (2019).
50. S. Chen, Y. Cai, G. Li, S. Zhang, and K. W. Cheah, "Geometric metasurface fork gratings for vortex-beam generation and manipulation," Laser Photon. Rev. **10**, 322–326 (2016).
51. G. Molina-Terriza, J. P. Torres, and L. Torner, "Twisted photons," Nat. Phys. **3**, 305–310 (2007).
52. K. Novak and A. T. Watnik, "Compact vortex wavefront coding camera," Proc. SPIE **11396**, 113960O (2020).
53. X. He, C. Liu, and J. Zhu, "Single-shot Fourier ptychography based on diffractive beam splitting," Opt. Lett. **43**, 214–217 (2018).
54. W. N. Plick and M. Krenn, "Physical meaning of the radial index of Laguerre–Gauss beams," Phys. Rev. A **92**, 063841 (2015).
55. S. Rasouli and D. Hebri, "Theory of diffraction of vortex beams from 2D orthogonal periodic structures and Talbot self-healing under vortex beam illumination," J. Opt. Soc. Am. A **36**, 800–808 (2019).
56. J. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Series in Electrical and Computer Engineering: Communications and Signal Processing (McGraw-Hill, 1996).
57. M. Guizar-Sicairos and J. R. Fienup, "Holography with extended reference by autocorrelation linear differential operation," Opt. Express **15**, 17592 (2007).
58. H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," arXiv:1708.07747 (2017).
59. T. Clanuwat, M. Bober-Irizar, A. Kitamoto, A. Lamb, K. Yamamoto, and D. Ha, "Deep learning for classical Japanese literature," arXiv:1812.01718 (2018).
60. H. M. Najadat, A. A. Alshboul, and A. F. Alabed, "Arabic handwritten characters recognition using convolutional neural network," in *10th International Conference on Information and Communication Systems (ICICS)* (IEEE, 2019).
61. Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010, http://yann.lecun.com/exdb/mnist/.
62. N. Baker, H. Lu, G. Erlikhman, and P. J. Kellman, "Deep convolutional networks do not classify based on global object shape," PLoS Comput. Biol. **14**, e1006613 (2018).
63. Z. Wang and A. Bovik, "Mean squared error: love it or leave it? A new look at signal fidelity measures," IEEE Signal Process. Mag. **26**(1), 98–117 (2009).
64. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science* (Springer, 2015), pp. 234–241.
65. K. Ikeuchi, ed., in *Computer Vision: A Reference Guide*, Springer Reference (Springer, 2014).
66. C. Zhu, E. A. Chan, Y. Wang, W. Peng, R. Guo, B. Zhang, C. Soci, and Y. Chong, "Image reconstruction through a multimode fiber with a simple neural network architecture," arXiv:2006.05708 (2020).
67. A. Greengard, Y. Y. Schechner, and R. Piestun, "Depth from diffracted rotation," Opt. Lett. **31**, 181–183 (2006).
68. S. R. P. Pavani and R. Piestun, "High-efficiency rotating point spread functions," Opt. Express **16**, 3484–3489 (2008).