

# Small-brain neural networks rapidly solve inverse problems with vortex Fourier encoders

Baurzhan Muminov and Luat T. Vuong\*

Department of Mechanical Engineering, University of California at Riverside, Riverside, CA, 92521, USA

\*Corresponding author: LuatV@UCR.edu

## ABSTRACT

We introduce a vortex phase transform with a lenslet-array to accompany shallow, dense, “small-brain” neural networks for high-speed and low-light imaging. Our single-shot ptychographic approach exploits the coherent diffraction, compact representation, and edge enhancement of Fourier-transformed spiral-phase gradients. With vortex spatial 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 this hybrid, optical-digital, vortex Fourier encoded, small-brain scheme, we reconstruct MNIST Fashion objects illuminated with low-light flux ( $5 \text{ nJ/cm}^2$ ) at a rate of several thousand frames per second on a 15 W central processing unit, two orders of magnitude faster than convolutional neural networks.

## 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.<sup>1</sup> It is fascinating how this area—iterative solutions of the Phase Problem—has developed and given rise to a set of optimization techniques that are today applied in other domains<sup>2</sup> and notably, provide the capacity to image through turbid and scattering media.<sup>3,4</sup>

Recently, it is also possible to obviate the Phase Problem for image construction with computational imaging. An expanding research area involves the application of neural networks, specifically deep-learning convolutional neural networks (CNNs).<sup>5</sup> 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.<sup>6–8</sup> When using coherent diffraction through phase masks, at least two distinct images are needed to attempt the Phase Problem.<sup>9–11</sup> The first (to the authors’ knowledge) application of CNNs for image reconstruction, is presented in,<sup>12</sup> where a phase-encoded image on a spatial light modulator is reconstructed via CNNs using intensity data from the camera. “Non-line-of-sight” CNN imaging is recently demonstrated from albedo autocorrelation patterns of speckles.<sup>13–15</sup> In these examples, as well as others that leverage neural networks, it is possible to reconstruct an *object type* without solving the Phase Problem, i.e., one may successfully predict an object using prior-trained patterns without being able to identify the *position* of the object.<sup>13</sup>

Still, deep learning neural networks offer additional functionality in the process of reconstructing the object. For example, simultaneous autofocusing with phase recovery<sup>16</sup> or super-resolution in pixel-limited or diffraction-limited systems.<sup>17</sup> 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.<sup>18</sup> The non-exhaustive list of important applications include profilometry,<sup>19</sup> imaging through smoke,<sup>20</sup> and LIDAR that leverages multiple point cloud and time-of-flight information.<sup>21</sup> Additional examples of “nonlinear reservoir learning” are

presented in,<sup>22</sup> which employs caustic patterns for original object reconstruction. The challenge with deep learning methods, however, is that the neural network requires large training sets, long training times, and these neural networks have higher degrees of computational complexity that render them vulnerable to adversarial network attacks.<sup>23</sup>

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.<sup>24,25</sup> 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.<sup>14,26–32</sup> What our scheme shows, unlike others, is that a small brain 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.<sup>33</sup>

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)<sup>34</sup> and robust (the resulting computer vision scheme is not susceptible to the rapid variations in scene illumination).<sup>35</sup> We note that, from a purely computational standpoint, Fourier representations have been demonstrated to be efficient at solving classification problems.<sup>36–38</sup> Object reconstruction with Fourier representations reduces requisite memory, power, or energy requirements and may even achieve real-time image processing.<sup>39,40</sup> The advantages of Fourier operations further multiply since they may be completed optically before the digital neural network.<sup>29</sup> Still, in each of the aforementioned cases using 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.<sup>41–47</sup> 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 is recently demonstrated in<sup>48</sup> where the reconstruction contrast ratio is increased as a result of the vortex phase.

In our approach, we achieve non-iterative, single-shot object reconstruction with a topological vortex-based lenslet-array design that contains multiple vortex phases in a lenslet pattern using the resulting edge-enhanced Fourier-plane representations. The presence of the vortex provides spatial encoding to break the translation invariance of the measured Fourier pattern and solve the Phase Problem. Image reconstruction is performed

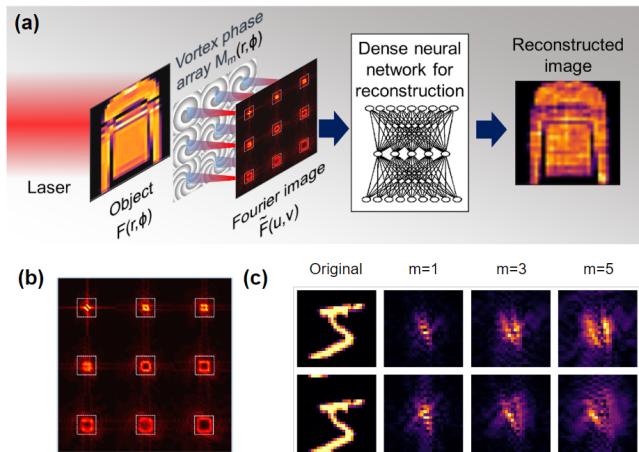


Figure 1. (a) The general schematic of the technique: a laser illuminates the object. Transmitted light is phase-modulated with a multi-vortex lens array. The back focal plane vortex-Fourier intensity patterns are fed to a neural net that reconstructs the original image. (b) The vortex-Fourier patterns have fewer pixels. Here, the combined area of 9 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  $m$ .

with 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".<sup>24</sup>

## 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  $M_m(r, \phi)$ ; and 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  $\phi$  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 [1(b)]. The vortex phase in the object 'real-domain' spatially encodes and breaks translational invariance of the Fourier-transformed intensity pattern 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:

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

where  $\alpha_0$  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  $\mathbf{X}$  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: 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.

### 2.1 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
- ptychographic mixing of real and imaginary parts to preserve phase in the intensity measurements.

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

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

where  $a$  is the radius of the mask aperture,  $\lambda$  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. At the phase plate [Eq. 2], the transmitted pattern is a sum of Laguerre-Gaussian modes at  $z = 0$  with different radial indices  $p$ ,

$$M_m(r, \phi) = \sum_p W_p LG_{m,p}(r, \phi), \quad (3)$$

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

$$LG_{m,p}(r, \phi) = L_p^{|m|} (2r^2/w^2) R(r) G(r) V_m(r, \phi) \quad (4)$$

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

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

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

where  $L_p^{|m|}(2r^2/w^2)$  are the generalized Laguerre polynomials that depend on  $m$  and  $p$ , and  $A_{m,p} = \sqrt{\frac{2^{|m|+1} p!}{\pi(p+|m|)!}} w^{-|m|}$ .<sup>49</sup> We expect that the waist of the beam  $w$  is larger than the features of the object  $F$ . The modal coefficients are,

$$W_p = \int \int M_m L G_{m,p}^*(r, \phi) r dr d\phi = A_{m,p} \int_0^a 2\pi r^{|m|+1} L_p^{|m|}(2r^2/w^2) G(r) dr. \quad (8)$$

The phase-singular term  $V_m$  is a radial magnitude gradient and azimuthal phase gradient, which can be simplified<sup>50</sup>

$$V_m(r, \phi) = A_{m,p} r^{|m|} e^{im\phi} = A_{m,p} [r \cos(\phi) + i \text{sgn}(m) \sin(\phi)]^{|m|}. \quad (9)$$

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  $\frac{v}{f\lambda}$ ,<sup>51</sup>

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

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,<sup>51</sup>

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

$$= \left| \sum_p \frac{W_p}{\lambda f} A_{m,p} \left( \lambda f \left[ \text{sgn}(m) \frac{\partial}{\partial v} - i \frac{\partial}{\partial u} \right] \right)^{|m|} \mathcal{F}\{F(r, \phi) L_p^{|m|}(2r^2/w^2) G(r)\} \right|^2 + \text{noise}. \quad (12)$$

In other words, the placement of the optical vortex provides unique optical preprocessing for the Fourier-plane data. Notably, the detected intensity patterns [Eq. 12] are composed of real and imaginary differentials of the Fourier transform of  $F$ . This differential scheme provides feature extraction via differentials in a manner similar to that deployed in the HERALDO method for image reconstruction.<sup>52</sup> At the same time, the vortex phase mixes the real and imaginary field components, to produce fringes in the intensity pattern.

We emphasize that the chirality in the system is critical.  $V_m$  represents a radial change in magnitude multiplied by an azimuthal change in phase. The resulting fringes exhibit a specific handedness associated with  $m$ , which preserves phase in the intensity patterns. For example, while the momentum-space representations of radial Laguerre polynomials<sup>49</sup> also provide spatial derivatives of Fourier components, they do not provide ptychographic encoding to solve the Phase Problem.

## 2.2 Small-brain machine learning of the inverse reconstruction mapping

Since a neural network is capable of guessing the reconstruction based on pre-learned patterns without solving the inverse or Phase Problem, we take a new approach towards understanding how the neural network learns these patterns. We test the inverse reconstruction of the neural network with several categorically-patterned datasets, namely the Fashion-MNIST,<sup>53</sup> Kuzushiji-MNIST<sup>54</sup> Arabic,<sup>55</sup> as well as the canonical handwritten MNIST digit dataset.<sup>56</sup> The “ground truth” outputs  $\mathbf{X}$  are the dataset’s 28x28-pixel images and are unit-normalized to provide comparable peak signal to noise (PSNR) with different image types across datasets. The vortex imaging scheme is capable of being applied as a “camera”; the vortex-based reconstruction achieves a mapping that is transferable or generalizable in cases when a random encoding scheme does not.

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, small-brain neural net with 1 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 simulations with numbers from the MNIST hand-written number dataset (flipped and inverted) [Fig. 2(a)]. However, we test with a separate dataset with patterns that are unrelated and different. With Kuzushiji characters [Fig. 2(b)], the reconstruction fails when there is no spatial encoding and also fails with a random spatial encoder [Fig. 2(c-d)], but succeeds, with some loss of resolution, with a vortex Fourier encoder ( $m = 3$ ) [Fig. 2(e)]. Without both compressed and encoded inputs and without previous patterns for guessing, the neural network cannot produce an inverse map from  $\mathbf{Y}$  to  $\mathbf{X}$ . With the Arabic data set [Fig 2(f-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 our unique opportunities with vortex-Fourier encoding schemes: a combination of compressed, encoded inputs is critical.

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. Fig. 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.

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.<sup>57</sup> Table 1 illustrates the convergence of the reconstructed images of the Fashion MNIST dataset to the original, both in terms of SSIM and MSE, as the number of vortices increases. We also employ the Structural Similarity Index Metrics (SSIM), which is limited to a  $[0,1]$  segment, which is a more reasonable metric for human quality perception evaluation.<sup>58</sup> For comparison, Table 1 also shows the SSIM and MSE for three-layer CNN-trained reconstruction with single and dual vortex datasets. This comparison suggests that our proposed architecture achieves the same quality while yielding much lower computational overhead (more than 3000 FPS

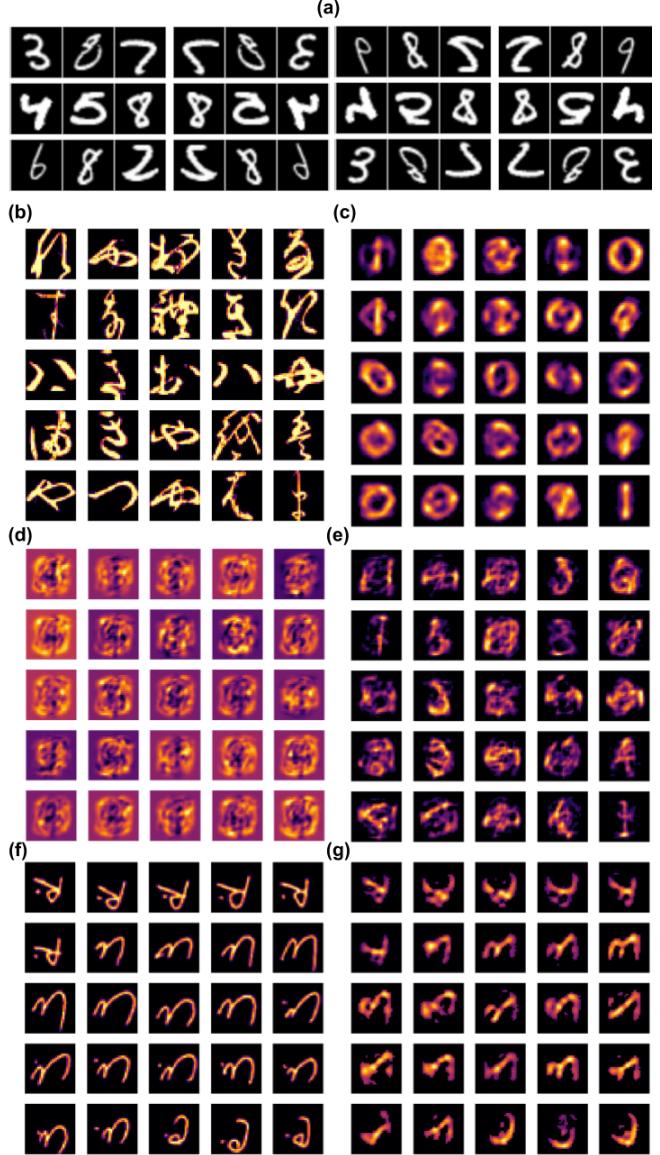


Figure 2. (a) 36 images of the training set composed of MNIST handwritten digit images flipped vertically and horizontally. (b) 25 images of the test set of MNIST Kuzushiji. Reconstructed images with (c) no spatial encoding, (d) random spatial encoding, and (e) vortex spatial encoding with  $m = 3$ . 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.

Table 1. SSIM and MSE for Fashion-MNIST reconstruction. The table shows that near-optimized quality is achieved with 2 vortices.

	Plain Fourier	1 vortex, linear	2 vortices, linear	3 vortices	CNN, 1 vortex	CNN, 2 vortices
SSIM	0.45	0.62	0.84	0.88	0.61	0.84
MSE	0.0280	0.0242	0.0140	0.0122	0.0235	0.0145

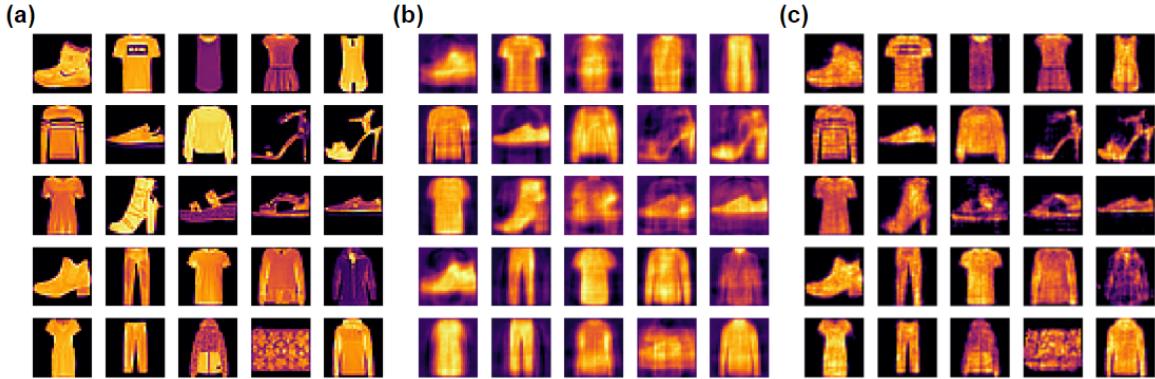


Figure 3. (a) 25 of the original Fashion-MNIST dataset images. Reconstruction with (b)  $m = 0$ , with no spatial encoding and (c) two vortices of topological charges  $m = 1, 3$  with non-linear activation is used in the last layer. Without spatial encoding, the neural network still learns the patterns when there are categorical variations between training and test sets. The structural similarity index metric (SSIM) is quantified in Table 1.

for the proposed network and less than 50 FPS for a three-layer CNN with 3x3 kernel).

### 2.3 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],<sup>59</sup>

$$noise = P_n(|\tilde{F}(u, v)|^2) + P_n(\sigma_d^2). \quad (13)$$

Both noise terms have Poisson distributions  $P_n(\mu)$ , where  $\mu$  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  $\sigma_d^2$  is related to the dark current and read noise of each pixel. The continuously-valued noise is related to the camera noise with MSE,

$$MSE = \frac{1}{N} \sum (y_0 - y_i)^2. \quad (14)$$

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 the our reconstruction algorithm. Therefore, regardless of the light intensity, we assume that the full dynamic range

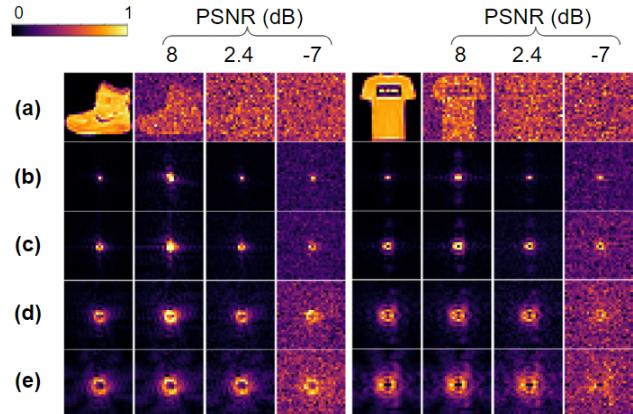


Figure 4. Normalized test inputs  $\mathbf{Y}$  to the neural network with noise for two Fashion-MNIST images. Each block shows decreasing peak signal to noise ratio (PSNR) in columns from left to right. (a) Object data  $\mathbf{X}$  and vortex spatial encoding (b)  $m = 0$  or no vortex (c)  $m = 1$  (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  $\mathbf{X}$ .

of a 12-bit camera is used. To vary the noise, we keep  $\sigma_d^2$  fixed and change  $|\tilde{F}(u, v)|^2$  to study the peak SNR (PSNR),<sup>58</sup>

$$PSNR = 10 \log_{10} \frac{P_{pk,signal}}{P_{noise}} \quad (15)$$

$$= 10 \log_{10} \left( \frac{(2^L - 1)^2}{MSE} \right), \quad (16)$$

$$= 10 \log_{10} \frac{\max(|\tilde{F}(u, v)|^2)}{\langle |\tilde{F}(u, v)|^2 \rangle + \sigma_d^2} \quad (17)$$

where  $P_{pk,signal}$  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). In other words, while we continuously vary the power to change the PSNR, we discretize the inputs to the neural network to use the maximum range of the 12-bit camera.

Figure 4 illustrates the tradeoff 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.<sup>22</sup> 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 Figure 5 with 2dB PSNR. Thus, vortex encoding provides feature extraction for efficient reconstruction of the object  $\mathbf{Y}$  to  $\mathbf{X}$ , 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 also tested with images from the same dataset with noise. (Again, only two vortices are needed to produce the best reconstructed image.) 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.

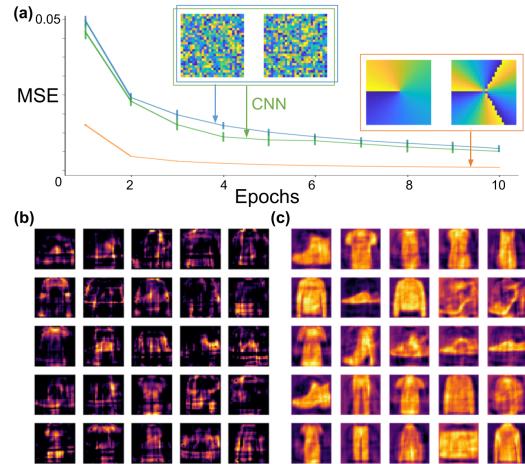


Figure 5. (a) Small-brain convergence (MSE vs. epochs) training the reconstruction of MNIST fashion images with dual vortex  $m = 1, 3$  (red) and random-pattern (blue) spatial encoders. The random encoder converges at a similar rate with a convolutional neural network (green). The reconstructed images with (b) random-pattern and (c) vortex encoders with 2dB PSNR.

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.

A comparison of different random patterns as spatial encoders suggests that the use of completely stochastic patterns are not best 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<sup>22</sup> is often applied to incoherent light illumination for image processing. There is certainly space for further study of topological structures for light scattering that connect different spatial-encoding and compressive-sensing approaches with neural networks.

### 3. EXPERIMENTS AND DISCUSSION

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: 1.1mm) 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: 1.5cm) onto the spatial light modulator (SLM)(Hammamatsu LCOS-SLM) at an incident angle of  $30^\circ$ . The SLM has  $800 \times 600$  pixels.

When recording experimental data, the phase patterns are saved in files at the start. An example phase pattern is shown in Fig. 7(b). Each pattern contains 6 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. 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  $4.65\mu\text{m} \times 4.65\mu\text{m}$  pixel area, 8-bit dynamic range, and  $1024 \times 768$  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  $30 \text{ nJ/cm}^2$  (intensities of  $10\mu\text{W}/\text{cm}^2$ , exposure times of 2.8 ms) over 6 vortex masks at the SLM. A “reference” CCD image that is taken with higher intensity is shown in Fig. 7(c). 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 of the camera read noise, even with minute-long camera exposure intervals.

With our automated matlab program, we record approximately 5 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 28x28 pixels for each  $\mathbf{Y}$ . A square area is cropped even though the CCD images are elongated. The validation/test, reconstructed, and SLM/CCD images taken with  $10\mu\text{W}$  power and 2.8-ms exposure times are shown in Figs. 7(d-g). All images are normalized to the color-range since

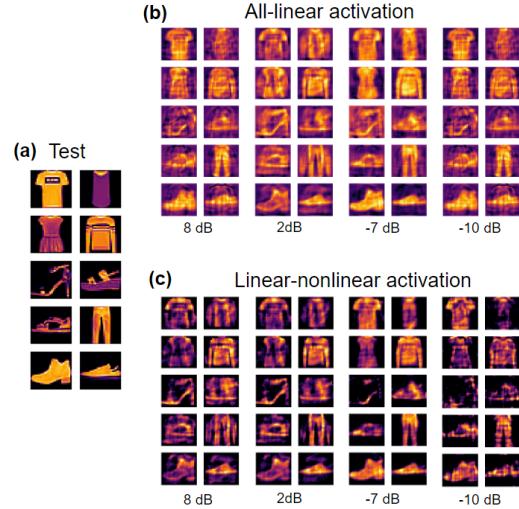


Figure 6. (a) Validation test images. Reconstruction results under PSNR-labeled noisy conditions that show the effect of (b) ‘linear’ and (c) ‘nonlinear’ activations for the final layer of the small-brain neural network. A vortex encoder with  $m = 1, 3$  is used.

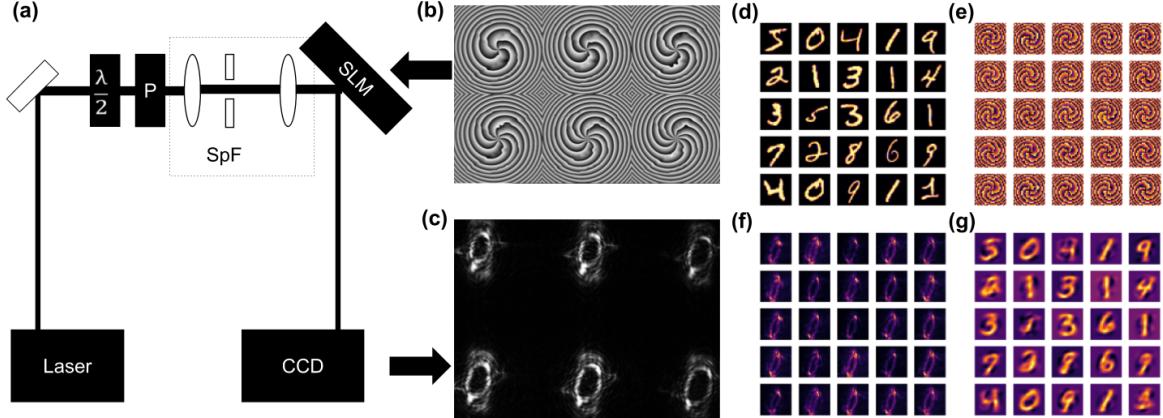


Figure 7. (a) Experimental setup. A laser ( $\lambda = 532$  nm) is directed through the half-wave plate ( $\lambda/2$ ), polarizer (P), and spatial filter (SpF) to obtain a near-perfect Gaussian beam. The lens focuses the beam on the spatial light modulator liquid crystal matrix (SLM) that operates in reflection mode. (b) The SLM pattern that imprints an  $m = 4 - 9$  vortex, a quadratic radial phase, and MNIST handwritten digit from the PC. The reflected light is collected on the CCD camera and stored on the same PC. (c) Example CCD image for vortices  $m = 4 - 9$  (d) 25 test images, (e) the corresponding phase patterns, (f) resulting CCD images and (g) reconstructed images.

without this colorbar normalization, the vortex Fourier patterns would not be visible. Even though there are 6 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. 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 neural network, 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 15W central processing unit, two orders of magnitude faster than convolutional neural net schemes. 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.

#### 4. CONCLUSION

We present a vortex-Fourier encoding approach to pre-processing 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 show that it is possible to solve inverse problems and obviate the need for CNNs 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 when random encoders do not. 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 specific spatial encoding by a vortex provides mixed edge-enhanced real and imaginary components, so that with 2 vortices, the neural network is transferable and generalizable. We aim to unbox the black box of machine learning by solving the generalized problem of inverse construction 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, resulting in images being less vulnerable to adversarial attacks.

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  $10\mu\text{W}/\text{cm}^2$  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 (10-20W).
- **Computationally efficient.** While other methods currently take multiple encoded images or use iterative schemes, we achieve near-to-ideal reconstruction with two ptychographic images.
- **Extremely low-power computation.** The technique uses explicitly simple neural nets (no deep learning) where pre-processing is completed with parallel optical propagation.
- **Compact with memory.** The vortex Fourier transform provides a compressed representation that can further be leveraged to minimize the number of pixels that carry data forward.
- **Flexible with a digital re-adjustable stage.** There is a tradeoff 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, 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. One envisions that the vortex-Fourier encoding scheme may be efficient at collecting radiation in pulsed, spectroscopic, laser applications. Further research may exploit topological features for achieving greater depth of field<sup>60,61</sup> and extend our knowledge of vortex Fourier ptychography to leverage information in light polarization, dispersion, and spatio-temporal coherence.

---

## REFERENCES

- [1] Latychevskaia, T., “Iterative phase retrieval in coherent diffractive imaging: practical issues,” *Applied Optics* **57**, 7187 (Aug. 2018).
- [2] Elser, V., Rankenburg, I., and Thibault, P., “Searching with iterated maps,” *Proceedings of the National Academy of Sciences* **104**, 418–423 (Jan. 2007).
- [3] Bertolotti, J., van Putten, E. G., Blum, C., Lagendijk, A., Vos, W. L., and Mosk, A. P., “Non-invasive imaging through opaque scattering layers,” *Nature* **491**, 232–234 (Nov. 2012).
- [4] Katz, O., Small, E., Guan, Y., and Silberberg, Y., “Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers,” *Optica* **1**, 170 (Sept. 2014).

- [5] Barbastathis, G., Ozcan, A., and Situ, G., “On the use of deep learning for computational imaging,” *Optica* **6**, 921 (July 2019).
- [6] Tahara, T., Quan, X., Otani, R., Takaki, Y., and Matoba, O., “Digital holography and its multidimensional imaging applications: a review,” *Microscopy* **67**, 55–67 (Feb. 2018).
- [7] Rivenson, Y., Wu, Y., and Ozcan, A., “Deep learning in holography and coherent imaging,” *Light: Science & Applications* **8** (Sept. 2019).
- [8] Konda, P. C., Loetgering, L., Zhou, K. C., Xu, S., Harvey, A. R., and Horstmeyer, R., “Fourier ptychography: current applications and future promises,” *Optics Express* **28**, 9603 (Mar. 2020).
- [9] Sidorenko, P. and Cohen, O., “Single-shot ptychography,” *Optica* **3**, 9 (Dec. 2015).
- [10] Zhang, F., Peterson, I., Vila-Comamala, J., Diaz, A., Berenguer, F., Bean, R., Chen, B., Menzel, A., Robinson, I. K., and Rodenburg, J. M., “Translation position determination in ptychographic coherent diffraction imaging,” *Optics Express* **21**, 13592 (May 2013).
- [11] Wang, H. and Piestun, R., “Azimuthal multiplexing 3d diffractive optics,” *Scientific Reports* **10** (Apr. 2020).
- [12] Sinha, A., Lee, J., Li, S., and Barbastathis, G., “Lensless computational imaging through deep learning,” *Optica* **4**, 1117 (Sept. 2017).
- [13] Metzler, C. A., Heide, F., Rangarajan, P., Balaji, M. M., Viswanath, A., Veeraraghavan, A., and Baraniuk, R. G., “Deep-inverse corereography: towards real-time high-resolution non-line-of-sight imaging,” *Optica* **7**, 63 (Jan. 2020).
- [14] O’Toole, M., Lindell, D. B., and Wetzstein, G., “Confocal non-line-of-sight imaging based on the light-cone transform,” *Nature* **555**, 338–341 (Mar. 2018).
- [15] Lei, X., He, L., Tan, Y., Wang, K. X., Wang, X., Du, Y., Fan, S., and Yu, Z., “Direct object recognition without line-of-sight using optical coherence,” in [2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)], IEEE (June 2019).
- [16] Wu, Y., Rivenson, Y., Zhang, Y., Wei, Z., Günaydin, H., Lin, X., and Ozcan, A., “Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery,” *Optica* **5**, 704 (May 2018).
- [17] Liu, T., de Haan, K., Rivenson, Y., Wei, Z., Zeng, X., Zhang, Y., and Ozcan, A., “Deep learning-based super-resolution in coherent imaging systems,” *Scientific Reports* **9** (Mar. 2019).
- [18] Li, Y., Xue, Y., and Tian, L., “Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media,” *Optica* **5**, 1181 (Sept. 2018).
- [19] Song, K., Hu, S., Wen, X., and Yan, Y., “Fast 3d shape measurement using fourier transform profilometry without phase unwrapping,” *Optics and Lasers in Engineering* **84**, 74–81 (Sept. 2016).
- [20] Locatelli, M., Pugliese, E., Paturzo, M., Bianco, V., Finizio, A., Pelagotti, A., Poggi, P., Miccio, L., Meucci, R., and Ferraro, P., “Imaging live humans through smoke and flames using far-infrared digital holography,” *Optics Express* **21**, 5379 (Feb. 2013).
- [21] Mitchell, E. W., Hoehler, M. S., Giorgetta, F. R., Hayden, T., Rieker, G. B., Newbury, N. R., and Baumann, E., “Coherent laser ranging for precision imaging through flames,” *Optica* **5**, 988 (Aug. 2018).
- [22] Antipa, N., Kuo, G., Heckel, R., Mildenhall, B., Bostan, E., Ng, R., and Waller, L., “DiffuserCam: lensless single-exposure 3d imaging,” *Optica* **5**, 1 (Dec. 2017).
- [23] Goodfellow, I., McDaniel, P., and Papernot, N., “Making machine learning robust against adversarial inputs,” *Communications of the ACM* **61**, 56–66 (June 2018).
- [24] Vuong, L. and Lane, H., “Nonlinear spectral preprocessing for small-brain machine learning,” in [Applications of Machine Learning], Zelinski, M. E., Taha, T. M., Howe, J., Awwal, A. A., and Iftekharuddin, K. M., eds., SPIE (Sept. 2019).
- [25] Horisaki, R., Takagi, R., and Tanida, J., “Learning-based imaging through scattering media,” *Optics Express* **24**, 13738 (June 2016).
- [26] Zhou, Y., Zheng, H., Kravchenko, I. I., and Valentine, J., “Flat optics for image differentiation,” *Nature Photonics* **14**, 316–323 (Feb. 2020).
- [27] Khoram, E., Chen, A., Liu, D., Ying, L., Wang, Q., Yuan, M., and Yu, Z., “Nanophotonic media for artificial neural inference,” *Photonics Research* **7**, 823 (July 2019).
- [28] Cordaro, A., Kwon, H., Sounas, D., Koenderink, A. F., Alù, A., and Polman, A., “High-index dielectric metasurfaces performing mathematical operations,” *Nano Letters* **19**, 8418–8423 (Nov. 2019).

- [29] Chang, J., Sitzmann, V., Dun, X., Heidrich, W., and Wetzstein, G., “Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification,” *Scientific Reports* **8** (Aug. 2018).
- [30] Zhu, T., Zhou, Y., Lou, Y., Ye, H., Qiu, M., Ruan, Z., and Fan, S., “Plasmonic computing of spatial differentiation,” *Nature Communications* **8** (May 2017).
- [31] Guo, C., Xiao, M., Minkov, M., Shi, Y., and Fan, S., “Photonic crystal slab laplace operator for image differentiation,” *Optica* **5**, 251 (Feb. 2018).
- [32] Lin, X., Rivenson, Y., Yardimci, N. T., Veli, M., Luo, Y., Jarrahi, M., and Ozcan, A., “All-optical machine learning using diffractive deep neural networks,” *Science* **361**, 1004–1008 (July 2018).
- [33] Muminov, B. and Vuong, L. T., “Fourier-plane vortex laser holography for robust, small-brain machine learning and image classification,” in [*Conference on Lasers and Electro-Optics*], OSA (2020).
- [34] Balboa, R. M. and Grzywacz, N. M., “Power spectra and distribution of contrasts of natural images from different habitats,” *Vision Research* **43**, 2527–2537 (Nov. 2003).
- [35] Krahmer, F. and Ward, R., “Stable and robust sampling strategies for compressive imaging,” *IEEE Transactions on Image Processing* **23**, 612–622 (Feb. 2014).
- [36] Rippel, O., Snoek, J., and Adams, R. P., “Spectral representations for convolutional neural networks,” in [*Advances in Neural Information Processing Systems 28 (NIPS 2015)*], Cortes, C and Lawrence, ND and Lee, D and Sugiyama, M and Garnett, R, ed., *Advances in Neural Information Processing Systems 28*, Neural Information Processing Systems (NIPS), 10010 NORTH TORREY PINES RD, LA JOLLA, CALIFORNIA 92037 USA (2015). 29th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, CANADA, DEC 07-12, 2015.
- [37] Alberti, M., Botros, A., Schuez, N., Ingold, R., Liwicki, M., and Seuret, M., “Trainable spectrally initialized matrix transformations in convolutional neural networks,” (2019).
- [38] Popa, C.-A. and Cernăzanu-Glăvan, C., “Fourier transform-based image classification using complex-valued convolutional neural networks,” in [*Advances in Neural Networks – ISNN 2018*], 300–309, Springer International Publishing (2018).
- [39] Chen, W., Wilson, J., Tyree, S., Weinberger, K. Q., and Chen, Y., “Compressing convolutional neural networks in the frequency domain,” in [*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*], ACM Press (2016).
- [40] Cherukara, M. J., Nashed, Y. S. G., and Harder, R. J., “Real-time coherent diffraction inversion using deep generative networks,” *Scientific Reports* **8** (Nov. 2018).
- [41] Wang, X., Nie, Z., Liang, Y., Wang, J., Li, T., and Jia, B., “Recent advances on optical vortex generation,” *Nanophotonics* **7**, 1533–1556 (Aug. 2018).
- [42] Allen, L., Beijersbergen, M. W., Spreeuw, R. J. C., and Woerdman, J. P., “Orbital angular momentum of light and the transformation of laguerre-gaussian laser modes,” *Physical Review A* **45**, 8185–8189 (June 1992).
- [43] Soskin, M. S., Gorshkov, V. N., Vasnetsov, M. V., Malos, J. T., and Heckenberg, N. R., “Topological charge and angular momentum of light beams carrying optical vortices,” *Physical Review A* **56**, 4064–4075 (Nov. 1997).
- [44] Lee, J. H., Foo, G., Johnson, E. G., and Swartzlander, G. A., “Experimental verification of an optical vortex coronagraph,” *Physical Review Letters* **97** (Aug. 2006).
- [45] Dennis, M. R., O’Holleran, K., and Padgett, M. J., “Chapter 5 singular optics: Optical vortices and polarization singularities,” in [*Progress in Optics*], 293–363, Elsevier (2009).
- [46] Zhu, L. and Wang, J., “A review of multiple optical vortices generation: methods and applications,” *Frontiers of Optoelectronics* **12**, 52–68 (Mar. 2019).
- [47] Chen, S., Cai, Y., Li, G., Zhang, S., and Cheah, K. W., “Geometric metasurface fork gratings for vortex-beam generation and manipulation,” *Laser & Photonics Reviews* **10**, 322–326 (Mar. 2016).
- [48] Novak, K. and Watnik, A. T., “Compact vortex wavefront coding camera,” in [*Computational Imaging V*], Petruccielli, J. C., Tian, L., and Preza, C., eds., SPIE (Apr. 2020).
- [49] Plick, W. N. and Krenn, M., “Physical meaning of the radial index of laguerre-gauss beams,” *Physical Review A* **92** (Dec. 2015).

- [50] Rasouli, S. and Hebri, D., “Theory of diffraction of vortex beams from 2d orthogonal periodic structures and talbot self-healing under vortex beam illumination,” *Journal of the Optical Society of America A* **36**, 800 (Apr. 2019).
- [51] Goodman, J., [*Introduction to Fourier Optics*], McGraw-Hill Series in Electrical and Computer Engineering: Communications and Signal Processing, McGraw-Hill (1996).
- [52] Guizar-Sicairos, M. and Fienup, J. R., “Holography with extended reference by autocorrelation linear differential operation,” *Optics Express* **15**(26), 17592 (2007).
- [53] Xiao, H., Rasul, K., and Vollgraf, R., “Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms,” (2017).
- [54] Clanuwat, T., Bober-Irizar, M., Kitamoto, A., Lamb, A., Yamamoto, K., and Ha, D., “Deep learning for classical Japanese literature,” (2018).
- [55] Najadat, H. M., Alshboul, A. A., and Alabed, A. F., “Arabic handwritten characters recognition using convolutional neural network,” in [*2019 10th International Conference on Information and Communication Systems (ICICS)*], IEEE (June 2019).
- [56] LeCun, Y. and Cortes, C., “MNIST handwritten digit database.” <http://yann.lecun.com/exdb/mnist/> (2010).
- [57] Baker, N., Lu, H., Erlikhman, G., and Kellman, P. J., “Deep convolutional networks do not classify based on global object shape,” *PLOS Computational Biology* **14**, e1006613 (Dec. 2018).
- [58] Wang, Z. and Bovik, A., “Mean squared error: Love it or leave it? a new look at signal fidelity measures,” *IEEE Signal Processing Magazine* **26**, 98–117 (Jan. 2009).
- [59] Ikeuchi, K., ed., [*Computer Vision: A Reference Guide*], Springer Reference, Springer, New York (2014).
- [60] Greengard, A., Schechner, Y. Y., and Piestun, R., “Depth from diffracted rotation,” *Optics Letters* **31**, 181 (Jan. 2006).
- [61] Pavani, S. R. P. and Piestun, R., “High-efficiency rotating point spread functions,” *Optics Express* **16**, 3484 (Feb. 2008).

## 5. ACKNOWLEDGEMENTS

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

## 6. FUNDING

The authors gratefully acknowledge DARPA YFA Grant #D19AP00036.

## 7. DISCLOSURES

The authors declare no conflicts of interest.