UC Riverside
UC Riverside Electronic Theses and Dissertations

Title

Optical Preprocessing for Low-Latency Machine Vision

Permalink

Author

Muminov, Baurzhan

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Optical Preprocessing for Low-Latency Machine Vision

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Mechanical Engineering

by

Baurzhan Muminov

September 2022

Dissertation Committee:

    Dr. Luat T. Vuong, Chairperson
    Dr. Nathaniel Gabor
    Dr. Erfan Nozari

The Dissertation of Baurzhan Muminov is approved:

_____

_____

_____
Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisor Prof. Luat T. Vuong, without whose help and patience, I would not have finished this Dissertation.

I would like to thank Prof. Nathaniel Gabor and Prof. Erfan Nozari for their emotional support and serving on my thesis committee despite their busy schedules.

I am grateful to my lab-mate Altai Perry for fruitful discussions and productive teamwork.

I would like to thank my collaborators Rakib Hyder and Prof. Salman Asif for useful discussions.

To my Mother.

ABSTRACT OF THE DISSERTATION

Optical Preprocessing for Low-Latency Machine Vision

by

Baurzhan Muminov

Doctor of Philosophy, Graduate Program in Mechanical Engineering
University of California, Riverside, September 2022
Dr. Luat T. Vuong, Chairperson

In recent years there has been an increased interest towards edge computing, *i.e.*, computing performed on distributed devices as opposed to centralized high-power hubs. Examples of edge computing would be the local image processing performed on Unmanned Autonomous Vehicles (UAV's) or the specialized machine vision systems on drones. These edge computing applications require schemes that are efficient with power and memory and typically must operate real-time. Many state-of-the-art image processing solutions that employ advanced optimization and deep neural networks (NNs) achieve impressive benchmark results, but are computationally demanding and thus on many occasions, impractical. The additional requirement for a range of applications is noise robustness or the ability to work in (extreme) low-light conditions; reasonable quality image or accurate object classification may be critical when there is low light flux or when the environment is over-saturated with other signals.

Here, we approach edge computing with a combination of optical preprocessing and shallow NN and we show that this hybrid approach greatly reduces the computational

requirements. For low-SNR imaging, we develop a technique that reconstructs objects and scenes from their Fourier-plane images. The optical preprocessing is performed via encoded diffraction with optical vortex singularities. The optical vortex encoder achieves differentiation of the already-compressed Fourier-plane patterns and enables facile inverse inference of the original object scene. We demonstrate that our method is robust to noise. And for a simple NN architecture (one or two layers), leads to generalization, *i.e.*, reconstruction of objects from classes that are greatly different from the ones the NN was trained on. Our research identifies strong potential for swift hybrid imaging systems with edge computing applications and highlights the valuable function of the vortex encoder for spectral differentiation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Simplifying and Speeding up the Solution of Phase Problem via Optical Computing

Our ability to solve inverse problems and reconstruct object features from either incomplete or mixed-signal components is essential for broad applications ranging 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: the sensor only registers intensity and therefore the phase information is lost. 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]. Iterative solvers of the Phase Problem have developed and given rise to a set of optimization techniques that are applied today in many other domains [2]. Notably, these

approaches provide the capacity to image through turbid and scattering media [3, 4, 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. One 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, 10, 11]. When using the coherent diffraction through phase masks, at least two distinct images are generally needed to attempt to solve the Phase Problem [12, 13, 14]. To our knowledge, the first application of CNNs for image reconstruction is presented in [15], where a phase-encoded image on a spatial light modulator is reconstructed via CNNs using intensity-only data from the camera. "Non-line-of-sight" CNN imaging is recently demonstrated from albedo autocorrelation patterns of speckles [16, 17, 18]. In these aforementioned 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].

In addition, deep learning neural networks offer extra 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 non-exhaustive list of important applications include: profilometry [22], imaging through smoke

or strongly-scattering media [23, 5], and LIDAR (Laser Imaging, Detection, and Ranging) 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].

1.2   Image Reconstruction via Universal Neural Networks

Image reconstruction has wide application in medicine [27, 14], biology [28], X-ray crystallography [29] and low-light vision, among other technologies. These reconstructions generally involve solving an inverse problem and retrieving the phase from phase-less intensity measurements. The field has been an active area of research for several decades [30, 31, 32] and inverse solvers achieve impressive results with additional coded optics or optical scanning [33, 25, 34, 35, 11, 36, 37, 38, 39, 40, 41, 42, 43, 44]. More recently, deep neural networks, and specifically convolutional neural networks, enable single feed-forward, non-iterative reconstruction [15] and are capable of learning from the statistical information contained in a variety of systems, from speckle [21, 45] to coded diffraction [46] patterns. Inverse solvers using neural networks are generally faster than iterative, optimization-based, or optical scanning-based algorithms and, as an example, may require as few as a 100 illumination training patterns with an "unrolled" neural network [47].
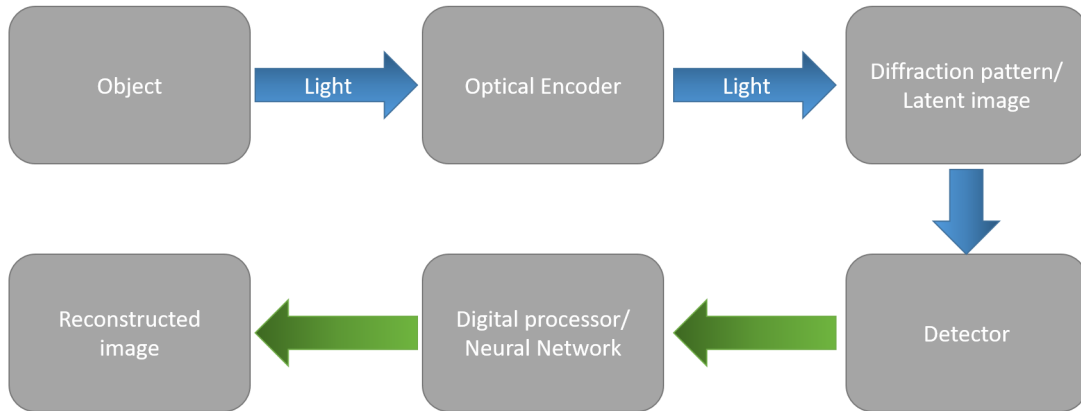
Figure 1.1: The general scheme used in this thesis. Light is reflected from the object, then passed through the (diffractive) optical encoder and collected by the camera (blue arrows). Then the data is processed electronically (green arrows). The last stage might be either reconstruction or classification.

Despite the benefits of neural networks to solve inverse problems, there are also drawbacks. Some of these issues—especially those associated with phase retrieval—have been solved. Other issues related to generalizability, robustness, and processing time or energy remain active areas of research [45]. Since neural networks learn how to weigh the importance of information patterns based on training data, they exhibit a tendency to "memorize" patterns to gain intuition about the task [48]. This predisposition towards prior data is advantageous for building "inductive, artificial intelligence machines" that extract patterns; however, that predisposition is a detriment to the generalizability of inverse solutions, (e.g., for building real-time computational cameras). Antun *et al.* [49] highlight three specific issues encountered by neural nets in imaging tasks:

1. Small, sometimes undetectable perturbations in the input (*both* image and sampling domain) can cause severe artifacts in the image reconstruction,

2. Small structural changes can be left undetected, and

3. More samples in the training set can lead to a deterioration of the results (due to the "memory" effect described above). Subsequently, algorithms themselves can stall or experience instabilities.

Designing simple, robust and computationally inexpensive ways to partially solve or alleviate these issues would be very useful for the field of inverse problems.

1.3   Machine Learning

In this thesis, we employ Machine Learning-based techniques to design simple inverse problem solvers which satisfy the requirements mentioned at the end of the previous section. This section serves as a brief, non-exhaustive description of Machine Learning.

Machine Learning is a field that builds a special class of algorithms which "learn" (are automatically tuned) via the use of data in the so-called "training" process. The results are prediction or decision making algorithms that, if used appropriately, can greatly outperform classical algorithms, especially where well structured information is not available. The canonical examples are: Computer Vision (object identification, image segmentation, movement prediction and other tasks [50]), and Speech Recognition (the translation of recorded voice as digitized acoustic waves into text [51]). Conventional Machine Learning techniques are split into two classes: 1) Supervised and 2) Unsupervised Machine Learning.

1.3.1 Supervised Machine Learning

With Supervised Machine Learning (SML) model is developed to produce a deterministic map from input data X to output data Y given some set of (X_0, Y_0) pairs of training data. Some examples may include prediction of: children's height given the age, race, sex, parents' height and other factors; the number of passengers in an airport given the day of week, time of the year, etc; the datacenter workload estimate given the noise level and spectral distribution created by its cooling system; crop yield in certain areas given the corresponding satellite images [52]. The free parameters (parameters that can be adjusted) of the corresponding model (map) are determined via minimization of the "loss function" that measures the discrepancy between the predicted and actual output data.

Overall, while being intuitively clear, SML typically requires large *marked* datasets. In some cases the collection of such a dataset might be rather tedious. However, for scenarios typically encountered in physics, marked dataset collection becomes straightforward using the automatic data collection tools. Let's consider the following examples: 1) predicting light intensity distribution given the incoming light distribution (for this problem we collect pairs of data that includes incoming light distribution and final light intensity distribution); 2) identifying the sequence of operations performed by the CPU given the electro-acoustic noise from the PC (analogously for this problem we collect the training data that includes sequence of operations performed by the CPU and electro-acoustic noise from the PC). For both of these cases we create a model with a number of variable parameters. The loss function likely would be different as in the first case we might use the "standard" measure e.g. Mean Squared Error between two light intensity distributions (predicted and

actual) and in the second it would be some discrete measure similar to Hamming distance between two sequences of operations (again: predicted and actual).

After the corresponding loss function is minimized, the model is considered "trained". Subsequently, the model is used to perform the predictions. However, the model's predictive performance must be evaluated on the data that is not used for training (validation dataset). The accuracy of the prediction on the validation dataset provides an estimate of the generalizability of the model. A model might be not useful if it "overfits" - an overly flexible model, with an excess number of parameters might memorize the patterns without the ability to generalize to previously unseen data. This can be tracked as loss function for training data keeps decreasing while loss function for validation dataset stays the same or even increases. Another common problem is adaptation to the specific data distribution, which limits the generalization capability of the resulting model. Ultimately, the task of choosing the appropriate and simultaneously efficient SML model architecture for solving problems in the physics domain is non-trivial. It requires understanding of the domain and there is no explicit procedure for choosing the optimal model architecture.

1.3.2   Unsupervised Machine Learning

Unlike SML, Unsupervised Machine Learning (USML) aims to identify patterns or structure in the data without any labels. The algorithm seeks commonalities in the provided sample characteristics/parameters and then clusters the data into distinct groups. Important examples are:

- Clustering - the process of separating data into different groups ("clusters") [53],

- Dimensionality reduction - the process of identifying "principal variables", to reduce the number of parameters for the multidimensional data. This is helpful both for visualization and processing purposes [54, 55],

- Anomaly detection - identifying the outliers in the data, e.g. identifying fraud operations in online banking [56, 57].

While USML possesses an attractive feature: it does not require marked datasets, it provides several challenges such as interpretability and, in general sense, "usefulness" for many physical applications. Overall, the USML is less intuitively clear than SML. In this thesis we did not explicitly use USML, however, it was employed to analyze the results and provide graphical representation of the data.

1.4 Informal Problem Statement

Machine Learning has found numerous successful applications in experimental image reconstruction tasks [8]. However it comes at a price: the current state-of-the-art solutions exploit very complex and computationally demanding neural networks. The one example of computationally costly algorithms are GAN's - Generative Adversarial Networks - algorithms in which two neural networks compete in generating and discriminating their predictions [58]. Such architectures, while providing impressive results, do not achieve real-time speeds and might take tens of hours on conventional hardware using hundreds of watts to reconstruct an image with several hundreds of pixels.

From one side, results as measured by different metrics (Mean Squared Error (MSE), Structural Similarity Index Metric (SSIM) [59], etc.) are impressive. From the

other side, architectural solutions are often chosen in order to achieve the highest possible scores in different metrics. This causes significant latencies (the delay of prediction after getting the data), especially in the case of edge computing (the computations done on the device capturing images or the device nearby) and require significant energy or power. Thus, the question of practicality arises; the problem is even more pronounced if we are concerned with the low power/low energy devices such as microdrones.

Given the previous paragraphs, one might pose a very interesting problem: can we simplify the networks to very simple architectures, while preserving a reasonable quality of reconstruction and classification? (We refer to this as a so-called "small-brain" problem). The answer is "yes". We can do that via very simple neural networks in conjunction with optical preprocessing done via various means. The resulting architectures demonstrate extremely low power and time requirements, while preserving a reasonable accuracy. Depending on the architecture, such imaging schemes might have the following advantages:

- Noise-robustness - the ability to operate in the extreme low-light mode (thousands of photons per picture; in comparison, typical "conventional" images require literally tens and hundreds of billions of photons to obtain images of reasonable quality). (This is due to the high-contrast ratio of the Fourier representation of the image). The preprocessing we perform allows us to achieve high quality reconstruction, which is troublesome if we were to use the "plain" Fourier imaging (i.e. unmodulated Fourier spectra),

- Interpretability - the predictions of the model can be "explained", the model does not perform as a black box. This property arises from the simplicity of the proposed

architectures - they can be analyzed via a simple Green function analogous process of single hot pixel analysis,

- Generalization - trained neural networks are able to reconstruct objects from previously unseen, significantly different classes of objects. We show that even such simple architectures possesses this property. To our knowledge this property for simplest neural networks is shown for the first time,

- Ability to be trained on the synthetic dataset (almost arbitrary set of images related to the transfer function of the optical system. Obviously, some initial "physical" samples are required to provide the rule for generation of the synthetic set), which almost eliminates the tedious task of data collection. In addition these synthetic datasets might be tuned for specific task, such as edge detection

In the next few chapters, we introduce and investigate the properties of several shallow neural networks preceded by the optical preprocessor. We demonstrate that our preprocessor in Fourier space performs feature extraction and thus greatly simplifies the process of image reconstruction. Also, the proposed architectures can be trained on synthetic dataset, thus reducing the amount of required training data. Resulting hybrid architectures, composed of fixed (optical) and variable (digital) parts, are noise-robust, extremely fast, and generalize on previously unseen data.

Chapter 2

Fourier Optical Processing for Low Light and Low Latency Machine Vision

2.1 Introduction

In this chapter, we focus on the application of shallow and fully-connected neural networks and ask, is it possible to achieve additional image-reconstruction functions without deep-learning and without iterative schemes? Such "small brain" approaches provide the advantage of a single forward pass, i.e., no iterative phase retrieval procedures [60, 61]. We demonstrate a simple approach to image reconstruction with optical preprocessing done by topological structures and lenses not employing the CNNs. Our strategy is similar to other hybrid and diffractive optical neural network approaches that aim to offload mathematical
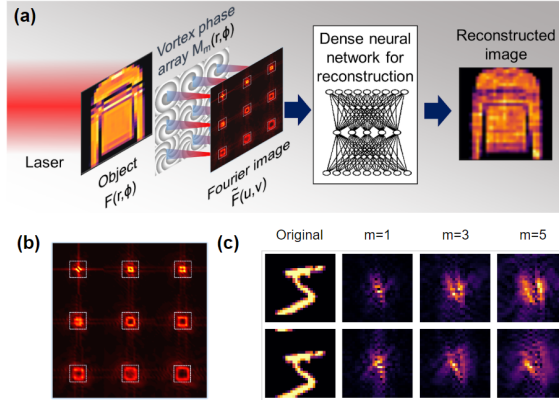
Figure 2.1: (a) The general schematic of the technique: a laser illuminates the object. Transmitted light is phase-modulated with a multi-vortex lens array. The back focal plane vortex-Fourier intensity patterns are fed to a neural net that reconstructs the original image. (b) The vortex-Fourier patterns 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.

computation to the propagation of light [62, 63, 64, 65, 66, 67, 68, 17], but aims to solve a specific problem of low-light/noise-robust imaging with extremely low latencies tuned for the edge processing done on the low-power/low-energy devices. 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 [69]. We do not make a claim that this is an optimal solution, but it's better than other more "classical" approaches and achieves the desired result. Currently there is no formal technique which can be use to retrieve the optimal prerpocessing for this task.

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) [70] and robust (the resulting computer vision scheme is not susceptible to the rapid variations in scene illumination) [71]. We note that, from a purely computational standpoint, Fourier representations have been demonstrated to be efficient at solving classification problems [72, 73, 74]. Object reconstruction with Fourier representations reduces memory, power, or energy requirements and may even achieve real-time image processing [75, 76]. The advantages of Fourier operations further multiply since they may be completed optically before the digital neural network [65]. 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 maps (modulations) in the form of 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 [77, 78, 79, 80, 81, 82, 83, 84]. 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 [85] where the reconstruction contrast ratio is increased as a result of the

13

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 [86].

In our approach, we achieve non-iterative, 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" [60].

2.2 Innovation

Figure 2.1 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 2.1(b)]. The vortex phase in the object 'real-domain' spatially encodes and breaks the translational invariance of the Fourier-transformed intensity pattern 2.1(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:

$$F(r, \phi) = e^{i\alpha_0 \mathbf{X}} \tag{2.1}$$

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 $\mathbf{X}$ 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.

2.2.1 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

- 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,

$$M_m(r, \phi) = \begin{cases} e^{\frac{-i\pi r^2}{\lambda f} + im\phi} & 0 \leq r < a \\ \\ 0 & otherwise \end{cases} \tag{2.2}$$

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 the Supplemental Content. ADD THE FIGURES FROM SUPPLEMENTAL INFORMATION! At the phase plate [Eq. 4.2], the transmitted pattern is a sum of Laguerre-Gaussian modes at $z = 0$ with different radial indices p,

$$M_m(r, \phi) = \sum_p W_p LG_{m,p}(r, \phi), \tag{2.3}$$

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

$$LG_{m,p}(r, \phi) = L_p^{|m|}(2r^2/w^2)R(r)G(r)V_m(r, \phi) \tag{2.4}$$

$$R(r) = e^{\frac{-i\pi r^2}{\lambda f}} \tag{2.5}$$

$$G(r) = \frac{1}{w}e^{-(\frac{r}{w})^2} \tag{2.6}$$

$$V_m(r, \phi) = A_{m,p}r^{|m|}e^{im\phi}, \tag{2.7}$$

where $L_p^{|m|}(2r^2/w^2)$ are the generalized Laguerre polynomials that depend on m and p, and $A_{m,p} = \sqrt{\frac{2^{|m|+1}p!}{\pi(p+|m|)!}}w^{-|m|}$ [87]. We expect that the waist of the beam w is larger than the

16

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,

$$W_p = \int \int M_m LG^*_{m,p}(r,\phi) r dr d\phi = A_{m,p} \int_0^a 2\pi r^{|m|+1} L_p^{|m|}(2r^2/w^2) G(r) dr. \qquad (2.8)$$

The phase-singular term V_m is a radial magnitude gradient and azimuthal phase gradient, which can be simplified [88]

$$V_m(r,\phi) = A_{m,p} r^{|m|} e^{im\phi} = A_{m,p}[r\cos(\phi) + ir\text{sign}(m)\sin(\phi)]^{|m|}. \qquad (2.9)$$

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 $\frac{v}{f\lambda}$ [89],

$$\tilde{V}_m(u,v) = \mathcal{F}\{V_m(r,\phi)\} = A_{m,p}\left(\lambda f\left[\text{sign}(m)\frac{\partial}{\partial v} - i\frac{\partial}{\partial u}\right]\right)^{|m|}, \qquad (2.10)$$

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 [89],

$$\mathbf{Y} = |\tilde{F}_m(u,v)|^2 + noise \tag{2.11}$$

$$= \left| \sum_p \frac{W_p}{\lambda f} A_{m,p} \left(\lambda f \left[\mathrm{sign}(m)\frac{\partial}{\partial v} - i\frac{\partial}{\partial u} \right] \right)^{|m|} \mathcal{F}\{F(r,\phi)L_p^{|m|}(2r^2/w^2)G(r)\} \right|^2 + noise \tag{2.12}$$

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

$$\nabla_\perp^2 V_m(r,\phi) = A_{m,p}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r} \right) + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} \right] r^{|m|}e^{im\phi} \tag{2.13}$$

$$= \left[\frac{|m|^2}{r^2} - \frac{m^2}{r^2} \right] V_m(r,\phi) \tag{2.14}$$

$$= 0. \tag{2.15}$$

Notably, the detected intensity patterns [Eq. 2.12] 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 [90].

2.2.2 Small-Brain Inverse Reconstruction

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
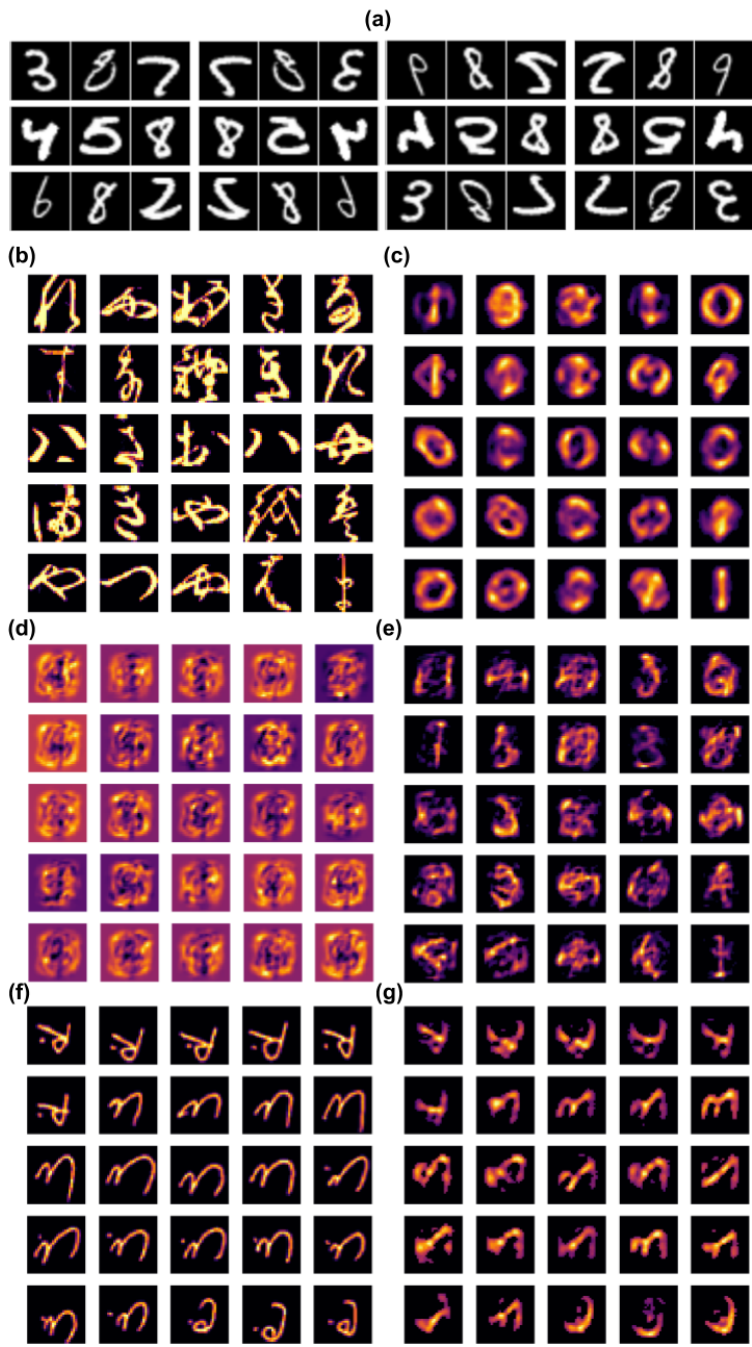
Figure 2.2: (a) 36 images of the training set composed of MNIST handwritten digits flipped vertically and horizontally. (b) 25 images of the MNIST Kuzushiji test set. Reconstructed images with (c) no spatial encoding, (d) random spatial encoding, and (e) vortex spatial encoding with $m = 3$. When we 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.

towards training and studying the neural network. We test the inverse reconstruction of the neural network with several categorically-patterned datasets, namely the Fashion-MNIST [91], Kuzushiji-MNIST [92], Arabic[93], as well as the canonical handwritten MNIST digit dataset [94]. 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-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. 2.12]. We set $\alpha_0 = \pi/2$ in a phase modulation scheme [Eq. 4.1]. 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 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. 2.2. We train the model with numbers from the MNIST hand-written number dataset (flipped and

inverted) [Fig. 2.2(a)]. However, we test with a separate dataset with patterns that are different. With Kuzushiji characters [Fig. 2.2(b)], the reconstruction fails when there is no spatial encoding and also fails with a random spatial encoder [Fig. 2.2(c-d)], but succeeds, with some loss of resolution, with a vortex Fourier encoder ($m = 3$) [Fig. 2.2(e)]. With the Arabic data set [Fig 2.2(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 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 Figure 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. Fig. 2.4(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 2.4(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 [95]. 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 [96]. This difference between SSIM and MSE illustrated Fig. S2 the Supplemental Content.

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 [97]). The reconstruction speed is measured using pre-collected data and we only consider the for the reconstruction: the
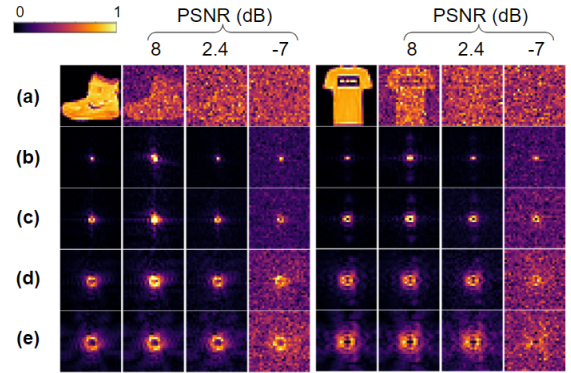
Figure 2.3: Normalized test inputs $\mathbf{Y}$ to the neural network with noise for two Fashion-MNIST images. Each block shows decreasing peak signal to noise ratio (PSNR) in columns from left to right. (a) Object data $\mathbf{X}$ and vortex spatial encoding (b) $m = 0$ or no vortex (c) $m = 1$ (d) $m = 3$ and (e) $m = 5$. The Fourier transform representations have higher PSNR that decreases with higher m given the same camera and light flux for $\mathbf{X}$.

Table 2.1: SSIM and MSE for Fashion-MNIST reconstruction with a small-brain (SB) and convolutional neural networks (CNN). U-Net CNN architecture is also implemented. Near-optimized quality is achieved with 2 vortices.

	SB, no vortex	SB, 1 vortex	SB, 2 vortices	SB, 3 vortices	CNN, no vortex	CNN, 1 vortex	CNN, 2 vortices (U-Net)
SSIM	0.45	0.62	0.84	0.88	0.41	0.61	0.84 (0.87)
MSE	0.0280	0.0242	0.0140	0.0122	0.0355	0.0235	0.0145 (0.0124)
Test	8600 FPS	8600 FPS	6400 FPS	5600 FPS	200 FPS	200 FPS	92 (27) FPS
Train	2 min	2 min	5 min	7 min	22 min	35 min	78 (440) min

CPU needed for Fourier transform is not in-

cluded 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 recon-

structs at a rate of more than 6000 images per second (FPS) while a three-layer CNN with

3x3 kernel achieves only 90 FPS. We also provide training s 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 4 and 5 vortices as inputs to the neural net-

work, we encounter increasing accuracy but diminishing returns. For example, 3 vortices

achieve SSIM = 0.88, 4 vortices SSIM = 0.89, 5 vortices SSIM = 0.91. The results for com-

parison using a simple CNN with a 3x3 kernel achieve similar SSIM with the same inputs.

Additionally, we implement a U-Net multiple-layer CNN [97] with 2 vortices. While addi-

tional 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.
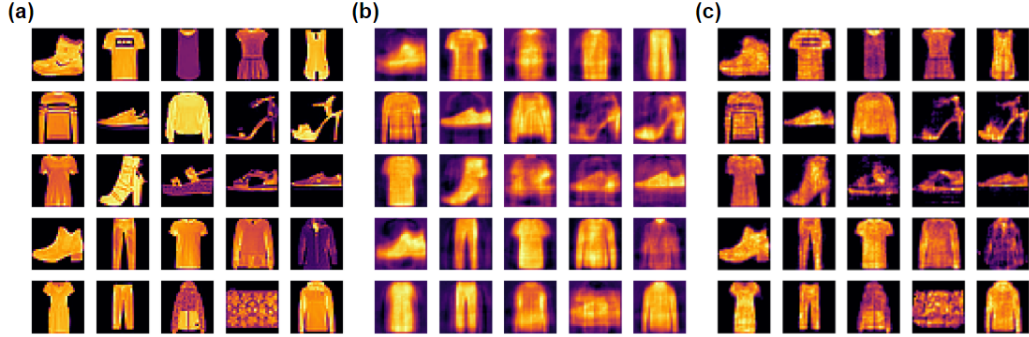
Figure 2.4: (a) 25 of the original Fashion-MNIST dataset images. Reconstruction with (b) $m = 0$, with no spatial encoding and (c) two vortices of topological charges $m = 1, 3$ with nonlinear activation used in the last layer. Without spatial encoding, the neural network still learns the patterns when there are categorical variations between training and test sets. The structural similarity index metric (SSIM) is quantified in Table 1.

2.2.3 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. 2.12] [98],

$$noise = P_n(|\tilde{F}(u,v)|^2) + P_n(\sigma_d^2).$$ (2.16)

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,

$$MSE = \frac{1}{N}\Sigma(y_0 - y_i)^2.$$ (2.17)
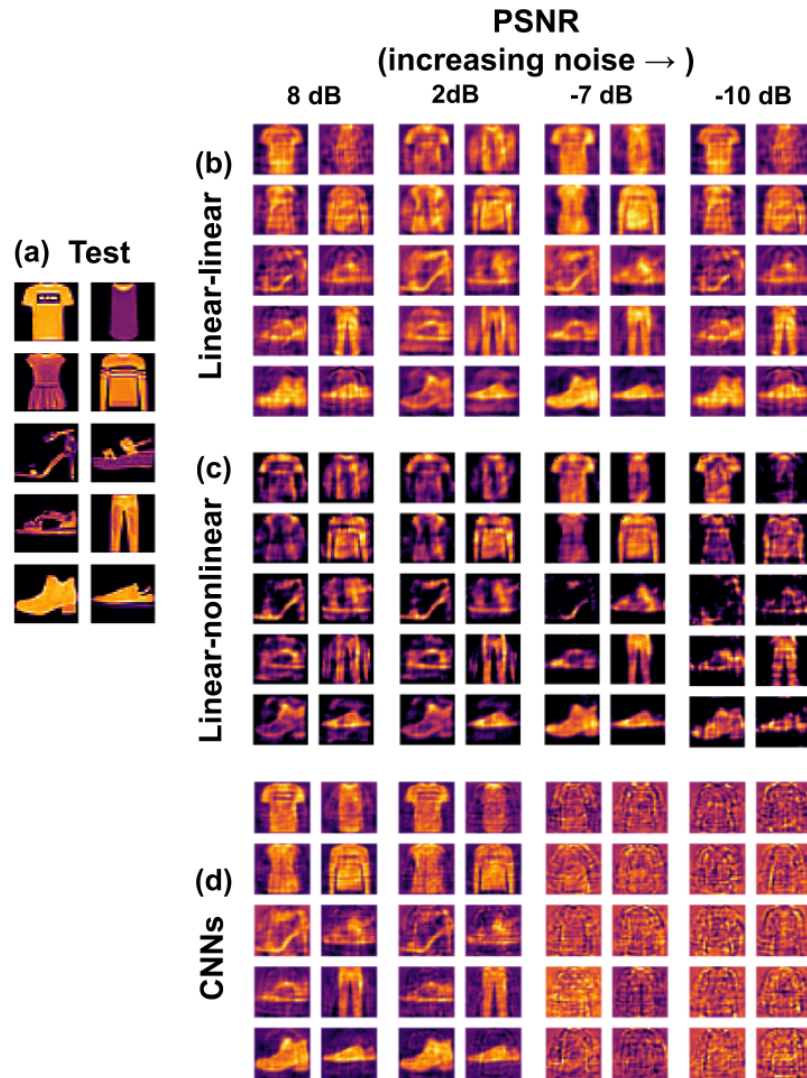
Figure 2.5: (a) Validation test images. Reconstruction results under PSNR-labeled noisy conditions that show the effect of (b) "linear" and (c) "nonlinear" activations for the final layer of the small brain or (d) convolutional neural network. A vortex encoder with $m = 1, 3$ is used.

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 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 peak SNR (PSNR) [96],

$$
\begin{aligned}
PSNR &= 10\log_{10}\frac{P_{pk,signal}}{P_{noise}} & (2.18)\\
&= 10\log_{10}\left(\frac{(2^L-1)^2}{MSE}\right), & (2.19)\\
&= 20\log_{10}\frac{max(|\tilde{F}(u,v)|^2)}{\langle|\tilde{F}(u,v)|^2\rangle+\sigma_d^2} & (2.20)
\end{aligned}
$$

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). The additional factor of 2 in Eq. 2.20 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 2.3 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 , when the area is larger, the effective PSNR decreases, resulting in a Fourier representation less robust to noise.

26

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 , 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 2.6 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 near-ideal reconstructed images.) The neural network learns an inverse mapping that minimizes MSE in the training set; however,
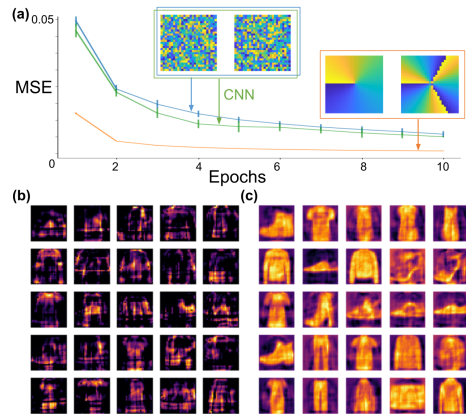
Figure 2.6: (a) Comparison of training rates (MSE vs. epochs) for different encoders in the reconstruction of MNIST fashion images. Dual-vortex encoding $m = 1, 3$ (red, small brain) converges faster than random-pattern encoding, which has a similar rate of convergence for a small brain (blue) and convolutional neural network (green). The reconstructed images with (b) random-pattern and (c) vortex encoders with 2dB PSNR.

different layer activations are vulnerable to

different types of noise. This is illustrated

in Fig. 2.5. 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.

2.3   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. 2.7(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 $TEM_{0,0}$ (Gaussian) beam is collimated (diameter: 1.5cm) onto the spatial light modulator (SLM)(Hammamatsu LCOS-SLM) at an incident angle of 30°. The SLM has 800600 pixels. Original images were upscaled in order to fill the SLM matrix size, so each spiral phase pattern is ˜165165 pixels. The neural network used for 2 vortex inputs is shown in Fig. 2.7(b).

When recording experimental data, the phase patterns are saved in files at the start. An example phase pattern is shown in Fig. 2.7(c). 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. The quadratic phase is an artificial lens for focusing and replaces a lens array, so we can collect all 6 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 4.65μm4.65-μm pixel area, 8-bit dynamic range, and 1024768 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 10μW/cm^2, exposure s 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. 2.7(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 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 , which is much longer than the camera shutter . The delay 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. The validation/test, reconstructed, and low-flux SLM/CCD images (10-μW power and 2.8-ms exposure) are shown in Figs. 2.7(e-h). All images are normalized to the color-range since 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. 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 increases by a factor of 4 compared to simulations. Longer training s 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 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
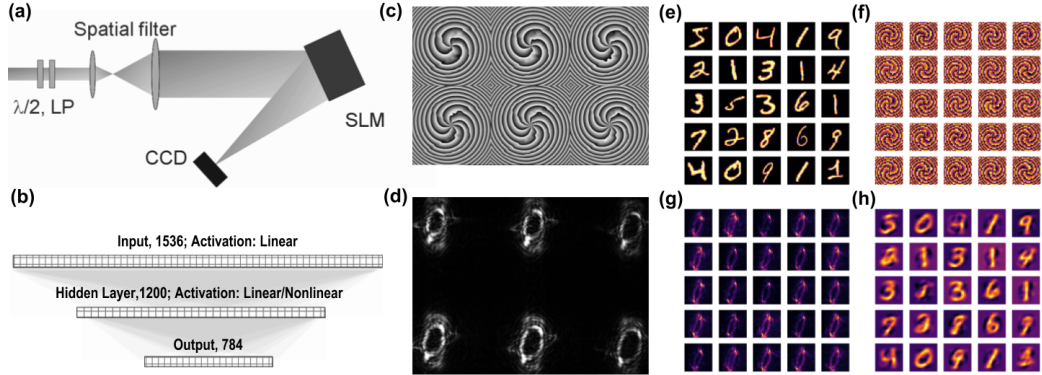
Figure 2.7: (a) Experimental setup. A laser ($\lambda = 532$ nm) is directed through the half-wave plate ($\lambda/2$), linear polarizer (LP), and spatial filter. The beam is collimated and illuminates the reflective spatial light modulator (SLM). A PC supplies patterns to the SLM. These vortex-object patterns also have a quadratic phase to focus all 6 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) 25 test images, (f) the corresponding 25 phase patterns with $m = 4$, (g) resulting CCD images with $m = 4$, and (h) reconstructed images.

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 [85]. 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 [1,2,3,4,5] 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 28x28 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 2 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 re-trained 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.

2.4   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 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. Given that more complicated neural network architectures can be unpredictable, [26], our approach that provides the 1-to-1 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 $10\mu\text{W}/cm^2$ average intensities with millisecond shutter s 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- 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 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 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. 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 multi-mode fiber was reported [99] 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, multi-mode fibers may provide another means for vortex encoding. Further research with free-space imaging may exploit topological features for achieving greater depth of field [100, 101] and extend our knowledge of vortex Fourier encoders to leverage information in light polarization, dispersion, and spatiotemporal coherence.

Chapter 3

Lightweight Universal Neural Network for Image Reconstruction

3.1 Introduction

Image reconstruction has wide application in medicine [27, 14], biology [28], X-ray crystallography [29] and low-light vision, among other technologies. These reconstructions generally involve solving an inverse problem and retrieving the phase from phase-less intensity measurements. The field has been an active area of research for several decades [30, 31, 32] and inverse solvers achieve impressive results with additional coded optics or optical scanning [33, 25, 34, 35, 11, 36, 37, 38, 39, 40, 41, 42, 43, 44]. More recently, deep neural networks, and specifically convolutional neural networks, enable single feed-forward, non-iterative reconstruction [15] and are capable of learning from the statistical information contained in a variety of systems, from speckle [21, 45] to coded diffraction [46] patterns.

Inverse solvers using neural networks are generally faster than iterative, optimization-based, or optical scanning-based algorithms and may require as few as a 100 illumination training patterns, for example, with an "unrolled" neural network [47].

However, despite the benefits of using neural networks to solve inverse problems, there are also drawbacks. Some of these issues—especially those associated with phase retrieval—have been solved. Others—particular those related to generalizability, robustness, and processing time or energy—remain active areas of research [45]. Since neural networks learn how to weigh the importance of information patterns based on training data, they exhibit a tendency to "memorize" patterns to gain intuition about the task [48]. This predisposition towards prior data is advantageous for building "inductive, artificial intelligence machines" that extract patterns; however, that predisposition is a detriment to the generalizability of inverse solutions, e.g., for building real-time computational cameras. Antun *et al.* [49] highlight three specific issues encountered by neural nets in imaging tasks:

1. Small, sometimes undetectable perturbations in the input (*both* image and sampling domain) can cause severe artifacts in the image reconstruction,

2. Small structural changes can be left undetected, and

3. More samples in the training set can lead to a deterioration of the results (as a result of the "memory" effect described above). Subsequently, algorithms themselves can stall or experience instabilities.

Whereas biomedical applications are aimed at large-image, high-quality image reconstruction [28], we turn our attention towards building real-time computational cameras for low size, weight, and power (SWaP) image reconstruction, which are needed for

autonomous-vehicle applications. In our prior effort [69], we demonstrate reconstruction with a "small brain" dual-layer neural network. Such regression-based approaches [61] demonstrate fast reconstruction rates, robustness to noise, and show potential for generalization with a phase vortex encoder. Here, we focus entirely on the generalizability of a simple neural network using a single-layer architecture for image reconstruction. We supply the model with a generalized or Universal Training Set (UTS) (synthetic images, used to train the neural network) and then test the neural network with images of different, unseen classes [see Fig. 3.1(a)]. A UTS-trained model overcomes the challenges associated with the "stereotypes" that generally arise from training by a specific image set. On the other hand, some disadvantages include the fact that the neural network is too simple to reconstruct images when nonlinear transformations are required [102]. Nevertheless, our results provide insight for training generalizable neural networks and computational cameras that operate at fast speeds. Our proposed method can readily be used for the initialization of alternating minimization problems or downstream image analysis tasks [103, 104, 105].

It is perhaps surprising that the simple learning model possesses enough capacity to recover a good approximation of the inverse coded-diffraction problem, and even with such a simple neural network there are interesting issues to address. In an effort to move towards producing a generalized training set, we compare the performance of the vortex encoder with other random encoders. From there, we build intuition for the UTS design based on the modal decomposition of the training, diffracted imaging patterns, and SVD-entropy. We also perform experiments, which build heuristics for real-world applications. We find that the choice of training images and optical encoder is important for achieving
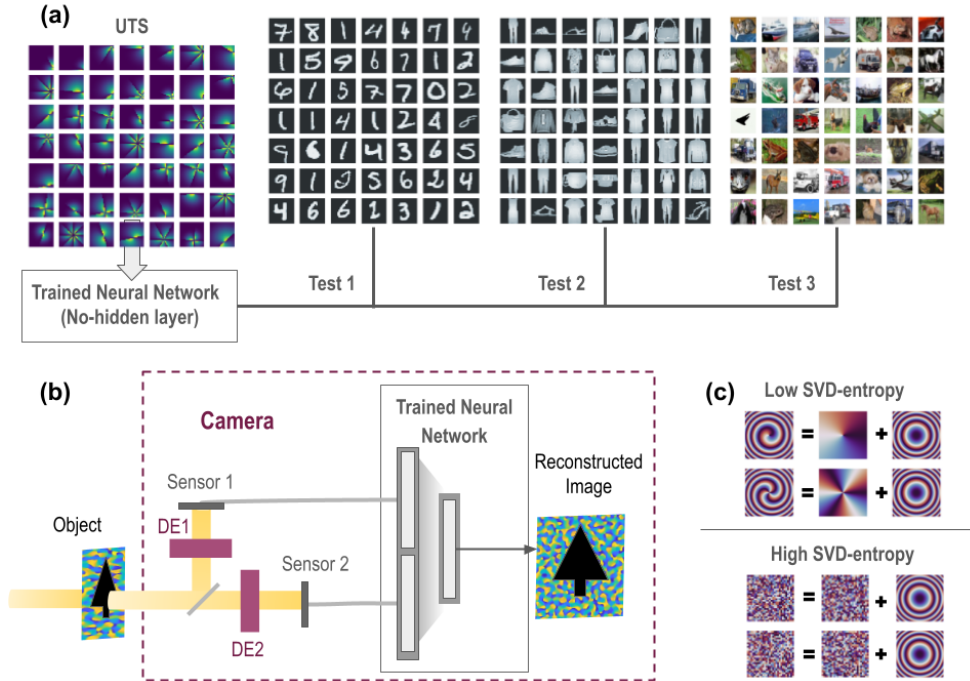
Figure 3.1: (a) Project objective: design a generalized training set for a neural network, which can later be used for general image reconstruction without retraining and operate real-time. (b) Schematic of hybrid vision camera where light from an object is transmitted through a diffractive encoder (DE). Sensors capture two transmitted images that are combined as inputs to the trained neural network, which reconstruct the object from the detector-plane images. (c) This project employs two pairs of diffractive encoders: one with low SVD-entropy (lens and topological charge $m = 1$ and 3) and the other with high SVD-entropy (uniformly-distributed random pattern).

generalizability, since not all imaged patterns provide a unique mapping to be learned and not all learned intensity patterns aid image reconstruction. While we have not quantitatively analyzed the image reconstruction *i.e.*, compared the set of training images to the span of the neural network, we observe that reduced SVD-entropy in the training set increases the learning efficiency, in both simulations and experiments.

3.2 Project Setup

In this section, we review an approach similar to [69] for our study of generalizable training. Figure 3.1(b) shows a schematic of the hybrid machine vision system, which encodes the image prior to the neural network with either a random or vortex phase pattern.

3.2.1 The Hybrid Vision System

The fields from the object at the diffractive encoder plane are $F(x, y)$. The encoder plane is imprinted with two diffractive element patterns $M(x, y)$, as shown in Fig. 3.1(c). A sensor or detector captures the intensity pattern of the electric fields $F'(u, v)$. Let $\mathcal{F}$ be the Fourier Transform operation $(x, y) \rightarrow (u, v)$, where we capture an image in the Fourier plane:

$$F'(u, v) = \mathcal{F}[M(x, y)F(x, y)] \tag{3.1}$$

Light from each object produces two images, each with a different diffractive element $M(x, y)$. Although the mask pattern may imprint vector (i.e., polarization-dependent) or spectral (time-dependent) delays, here we assume a homogeneous polarization, a linear encoder, and monochromatic, continuous-wave light. All optical neural networks have been previously demonstrated, notably with several diffractive layers in the THz regime [106], with nonlinear activations via saturable-absorbing nonlinearities [63], and with nano-interferometric etalons [107] in the visible regime. All-optical methods maximize speed and minimize energy loss in the neural computation [108]. At the same time, all-optical systems require nonlinear interactions as proxies for the electronic neural network layer activations. These nonlinearities occur at small length scales in order to confine light sufficiently, so

all-optical computing may be more sensitive to environmental conditions and less suitable for autonomous-vehicle computational cameras.

By contrast, we focus on hybrid imaging in which optical processing conditions sensor measurements and an electronic neural network performs reconstruction [109, 110, 65]. Our work is also inspired by pytchography approaches in [34, 35, 11]. Two phase masks are used to capture the intensity measurements of the object on the sensor, which are then fed to a no-hidden-layer neural network. At this time, we do not predict depth sensing with imaging, so the masks contain lenses for Fourier-plane detection. Here we reproduce the object based on the detector intensity patterns and assume that the detector is in the focal plane associated with a quadratic radial phase of the mask. In recent work, Fresnel mid-field imaging shows potential for better object-based depth detection [111].

In a manner similar to [69], we generate phase-modulated patterns,

$$F(x, y)M(x, y) = e^{i\alpha \mathbf{X}} G(x, y)M(x, y), \tag{3.2}$$

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

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

$$\mathbf{Y} = H(\mathbf{X}) + \mathbf{N}, \tag{3.3}$$

or for our specific case,

$$\mathbf{Y} = |\mathcal{F}[e^{i\alpha\mathbf{X}}G(x,y)M(x,y)]|^2 + \mathbf{N}, \tag{3.4}$$

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

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

3.2.2 Universal Training Sets (UTS) and Diffractive Encoders

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

$$M(x,y) = e^{-(x^2+y^2)(\frac{i}{f\lambda}+\frac{1}{w^2})}e^{im\phi} \tag{3.5}$$

where f is the effective focal length of the radial quadratic phase, λ is the wavelength of light, m is an on-axis topological charge, and w is the width of the Gaussian beam illuminating the mask. Figures 3.1(b,c) show diffractive elements with $m = 1, 3$. The second pair is composed of random masks, where each pixel of the transmitted pattern is encoded with a

Figure 3.2: Reconstructed images from (a, b, c) MNIST handwritten and (d, e, f) fashion MNIST datasets with random, Fourier and vortex bases, respectively. The vortex basis provides edge enhancement for object detection. (g) Ground truth and (h) reconstructed images from the CIFAR-10 dataset using the vortex training bases and a vortex mask as the encoder.

random phase from 0 to 2π. The mask is also illuminated with the same Gaussian beam. On the side of the training, we work with a range of images composed of 28×28 patterns that are random $\mathbf{X}_R$, Fourier-based $\mathbf{X}_F$, or shapes related to a vortex phase $\mathbf{X}_V$.

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

$$\mathbf{X}_{F(s_j, s_k, \phi_l, n)}(x, y) = \angle(e^{i(xs_j + ys_k + \phi_l)})G_n \tag{3.6}$$

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

$$G_n(x, y) = e^{-[(x-x_n)^2 + (y-y_n)^2]/w_n^2} \tag{3.7}$$

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

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

$$\mathbf{X}_{V(x_j,y_k,\phi_j,n,l)}(x,y) = \angle\{e^{im_l \tan^{-1}[(y-y_k)/(x-x_j)]+\phi_k})G_{j,k,n}. \tag{3.8}$$

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

$$G_{j,k,n}(x,y) = e^{-[(x-x_j)^2+(y-y_k)^2]/w_n^2}. \tag{3.9}$$

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

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

Once trained with a large dataset, we observe that the dense neural network without hidden layers can approximate almost any shape-based image (MNIST, fashion MNIST, CIFAR). An example set of reconstructed images from different classes is shown in Fig. 3.2. Figure 3.2 shows a representative set of images reconstructed from models trained with $\mathbf{X}_F, \mathbf{X}_V$, and $\mathbf{X}_R$ and a vortex mask. In each case, 20,000 training images are used. Error with thresholding is as low as 10% with test datasets. While the overall error is similar, models trained with the vortex-phase datasets, $\mathbf{X}_V$, generally have the lowest error and

strongly highlighted edges. Meanwhile models trained with a Fourier basis $\mathbf{X}_F$ have the highest error and models trained with a random basis $\mathbf{X}_R$ have error in between, with error distributed over the area of the image. Additional differences are explained in the following section.

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

With this simple neural network and three different UTS, we observe trends in convergence and overfitting. These trends consistently depend on the choice of the UTS patterns regardless of the choice of mask M_V or M_R. Figures 3.3(a-c) show samples from 20k-image $\mathbf{X}_F, \mathbf{X}_V,$ and $\mathbf{X}_R$ UTS with the vortex mask M_V. Some pairings converge with minimal overfitting while others do not provide enough information in $\mathbf{Y}$ to calculate the inverse of the nonlinear mapping, $H(\mathbf{X})$ [Fig. 3.3(d-f)].

A Fourier basis is the most well-known spectral basis for decomposing an image. When training with a Fourier basis, the validation loss stops decreasing after a certain number of epochs, which signals that the neural network struggles to extract information about the mapping given this orthogonal set of images. What this tells us is rather unintuitive about the span or basis of image reconstruction with neural networks, but potentially addressed in [112]: the images are less effectively learned by the neural net because there is minimal overlap between them; the correlations between Fourier modes are less visible to the neural net.
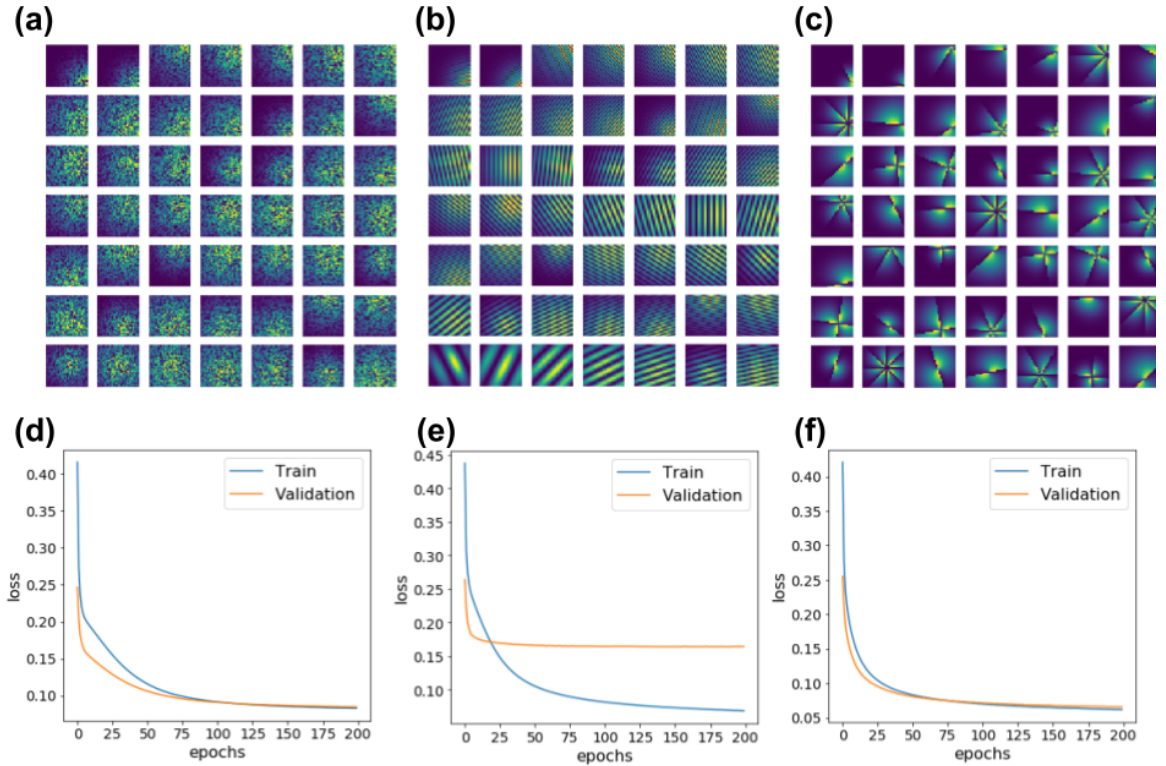
Figure 3.3: (a,b,c) Sample training images $\mathbf{X}_R$, $\mathbf{X}_F$, and $\mathbf{X}_V$ or random, Fourier, and vortex training sets. (d, e, f) Corresponding training and validation curves.

The random UTS also unreliably converges when the dataset is smaller than 2k, and its loss generally shows a "hill", where the loss plateaus before dropping. Meanwhile, the vortex-based UTS is less prone to such behavior. This combination of trends tells us that neither orthogonality nor randomness is ideal for training a neural network. The structured pattern of our vortex-based UTS $\mathbf{X}_V$ is a better candidate for generalized training compared to random $\mathbf{X}_R$ or Fourier $\mathbf{X}_F$ patterns. In our discussion, we provide some measures related to the UTS image analysis and trained model robustness.

3.3   Discussion

In this section, we discuss the ability to recreate sharp images, which may be seen by the single-pixel response. The single-pixel response from the random UTS-trained neural network is sharply corrugated [Fig. 3.4(a)], whereas the structured, single-pixel images from the vortex-trained model is generally smooth with a sharp "hole" in the center or dark spot [Fig. 3.4(b)]. We claim that these differences in the impulse response are responsible for the edge-enhanced reconstruction of shapes in Figs. 3.2(c,f). Figures 3.4(a,b) illustrate example images reconstructed with just one "hot" pixel in the camera sensor plane. These patterns are the building blocks of the reconstruction scheme and these patterns change depending on how the model is trained. Depending on the training set, the model is tuned to pay attention to different features of the image, which may depend on the task at hand.
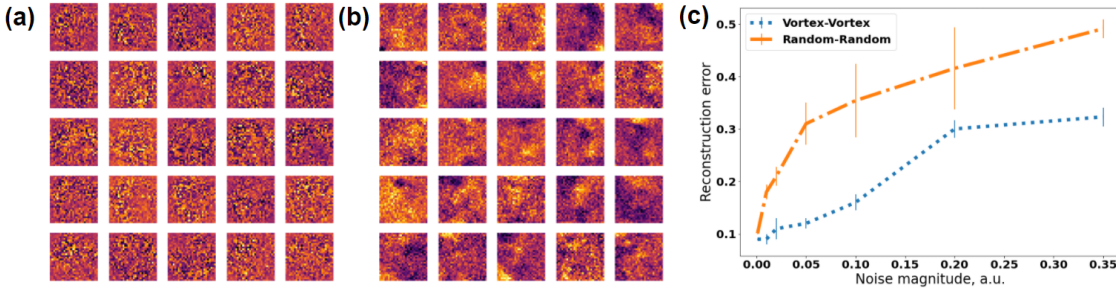
Figure 3.4: (a) Single "hot" pixel response of the random model and (b) single-pixel response of vortex model, which demonstrates sharp edges and resolves high-contrast objects. (c) Comparison of reconstruction error for different levels of noise given high-entropy random UTS and random mask and lower SVD-entropy vortex UTS and vortex mask. This error corresponds to the scenario in which shot noise dominates the background noise.

Figure 3.4(c) provides a simple noise analysis that shows the additional advantage of robustness when the neural network is trained with a low-entropy UTS. We show the

reconstruction error as a function of noise magnitude. Poisson shot noise and background noise are added to the Fourier-plane intensity patterns of the test image set. Low SVD-entropy image training and encoders appear more robust.

3.3.1   Analysis with Singular Value Decomposition Entropy

In order to estimate complexity of the pattern we employ the measure of entropy. We approximate the 2D entropy of the images using the spectra of singular value decomposition (SVD), which describes the complexity of an image. Unlike Shannon entropy [113], SVD-entropy illustrates the mixture of spatial modes that are present in an image.
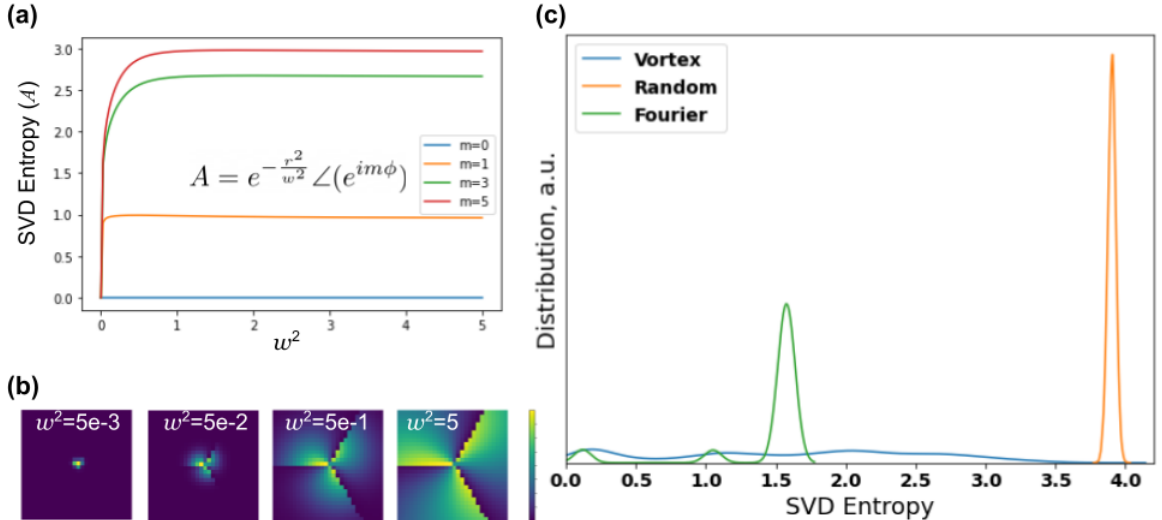
Figure 3.5: (a) SVD-entropy of a structured pattern composed of the phase of a vortex (modulus 0, 2π) and a Gaussian mask with radius of w^2. A few-pixel pattern has almost zero entropy, and the SVD-entropy saturates for a vortex depending on the topological charge. (b) Illustration of these patterns with $w^2 = [5e - 3, 5e - 2, 5e - 1, 5]$ corresponding to SVD-entropy values of [0.94, 1.8, 2.6, and 2.7]. The SVD-entropy strongly relates to the length of the edge dislocations of an image. (c) Histogram of the SVD-entropy in the vortex $\mathbf{X}_V$, Fourier $\mathbf{X}_F$, and random $\mathbf{X}_R$ generalized training sets implemented in this project.

51

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

$$E_{SVD} = -\sum_{1}^{K} \overline{\sigma_i} \log_2(\overline{\sigma_i}), \tag{3.10}$$

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

$$\overline{\sigma_i} = \frac{\sigma_i}{\sum_{1}^{K} \sigma_i} \quad \text{and} \quad \sum_{i} \overline{\sigma_i} = 1, \tag{3.11}$$

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

Some trends related to the SVD-entropy are illustrated in Fig. 3.5. If images in the set have several high singular values σ_i, the images may be reconstructed using fewer "elementary" patterns; those with higher entropy require many more patterns to achieve enough reconstruction accuracy. Low SVD-entropy images are smoother with fewer edges. On the other hand, images with many discontinuities exhibit a high degree of SVD-entropy.

From our analysis of differently structured patterns, the SVD-entropy scales logarithmically with the edge steps or dislocations in an image [Fig. 3.5(a-b)]. In this illustration, we plot the phase of an $m = 3$ vortex with varied Gaussian-beam filtering. The measure of 2D SVD-entropy aids our analysis of the UTS. The vortex UTS has a broad range and lower values of SVD-entropy in contrast to the random UTS [Fig. 3.5(c)].

Pertaining to our efforts towards generalized training or a UTS, we see that a low SVD-entropy training set like that with structured patterns $\mathbf{X}_V$ allows us to extract the structured (low SVD-entropy) information from the data [Fig. 3.2 (c,f,g,h)]. This effectively

acts as a filter for salient features of the image. This low SVD-entropy training would be useful for some specific tasks, especially when, e.g., we are less interested in the image's background information than in the foreground object.

3.3.2 Heuristic Experiments
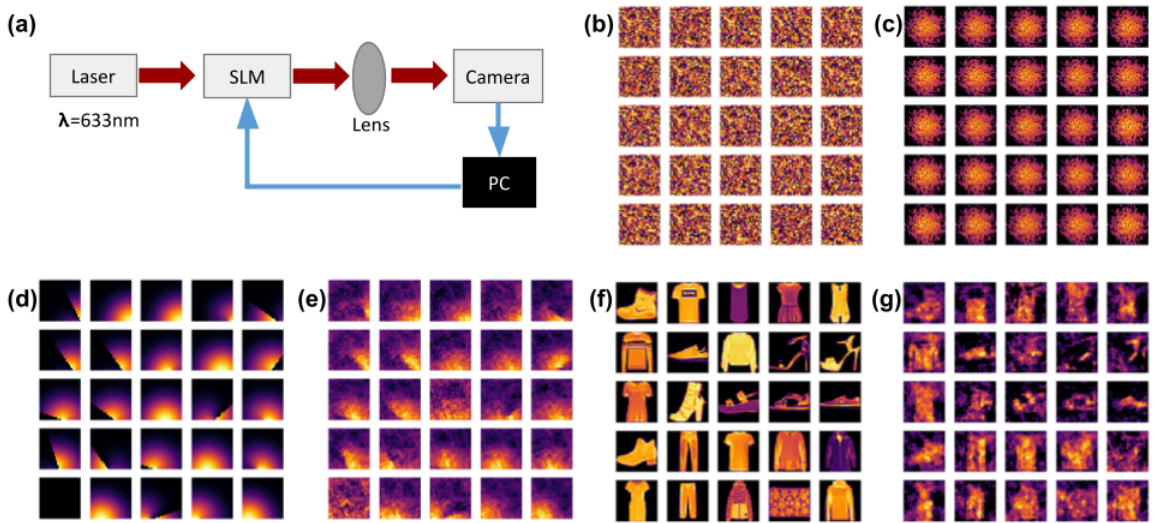
Figure 3.6: (a) Schematic of experimental reconstruction with UTS. There is no spatial filter or polarizer, images are noisy and at this wavelength, the modulation dynamic range is only $\alpha = \pi$. This was done intentionally to simulate poor experimental conditions with background light. (b) Sample random UTS images and (c) sample reconstructed images produced by random patterns, which are not learned by the simple neural network model experimentally. On the other hand, (d) simpler images with fewer edges are (e) reconstructed by the neural network. (f) Sample ground truth images and (g) discernable reconstructed patterns when the neural network is trained by the vortex dataset.

To illustrate the potential and the impact of our approach for generalizable training, we show heuristic experimental results. In simulations, almost any encoded diffraction pattern with a mask presents a learnable map for a simple neural network. However, in practice when noise is present, neural networks do not always converge. Our experimen-

tal data show that under noisy experimental conditions where light is unpolarized and the sensor data is collected with significant levels of noise, the high SVD-entropy dataset is not suitable for the task of image reconstruction: background light and distortions render a high SVD-entropy training image useless since the neural network does not learn the pattern. By contrast, a neural network trained on low SVD-entropy images is capable of recovering reasonable approximations of the unseen images, as shown in Fig. 3.6.

Our experimental setup consists of a 633-nm Helium-Neon continuous-wave laser, microscope objective, HOLOEYE Spatial Light Modulator and focusing lenses, and a CMOS 8-bit camera (1280x1024 pixel resolution). The setup does not include polarizers as part of the design to provide a large-background and an unmodulated signal to test the limits of image reconstruction with a simple neural network. As a result, we are unable to recover images with the zeroth-order transmitted pattern. When we instead collect the sensor data at the first diffraction maximum, we are successful with image reconstruction but only with the vortex UTS. For reconstruction purposes, small square patches of the detector pattern are taken (e.g. 50×50 pixels).

In our experiments with imperfect spatial beam profiles and background unmodulated noise, the simple neural networks do not converge with random masks (the results are shown in Fig. 3.6(b-c)). Experimentally, we demonstrate two masks shown in Fig. 3.1(c-d]), which are successfully learned by the neural network. The low SVD-entropy dataset composed of shapes with straight edges and curves, i.e., $\mathbf{X}_V$ [Eq. 3.8] converges but the high SVD-entropy random $\mathbf{X}_R$ patterns do not. Again, we find it more difficult to train a simple neural network with a high SVD-entropy UTS.

3.4   Conclusion

Corners, edges, and higher-order solutions are a challenge in image reconstruction, requiring a higher degree of superposed waves [114]. This more complex representation of images is the definition of SVD-entropy in an image, and suggests that the reconstruction of such images requires the learning of images composed of high SVD-entropy patterns [115]. We find, however, that this is not always the case when aiming for robust neural network-based reconstruction. In fact, generalized training with low-entropy patterns recreates these sharp features well with edge enhancement.

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

We have previously shown that a convolutional neural network can outperform a single layer neural network but with significantly higher energy cost. The deep neural network is also less robust to noise [69]. Here, we aim to work with a "small brain" neural network rather than a deep neural network architecture. This approach has been specifically tuned with the aim of low-SWaP computational cameras. We conclude that

- Single-layer neural networks are capable of approximating the inverse mapping from phaseless Fourier-plane intensity patterns after basic training.

- Such moderate-accuracy generalizable image reconstruction achieves high speeds (we achieve 15k fps on a 15W laptop CPU).

- Image reconstruction with simpler neural networks are robust to the vulnerabilities and instabilities described by [49].

- Even with a simple neural network architecture and a large training basis set, we encounter differences in convergence. (Experimentally with an imperfect encoder, neural networks learn low SVD-entropy images more rapidly and reliably than high SVD-entropy.)

- Low SVD-entropy images are valuable in training neural networks to extract the salient features of the image.

Additional advantages of a UTS include what is likely a generalized upper bound for error [28], higher robustness, and high potential for low-SWaP computational cameras. Because of its low computational complexity, our approach in the future may be inverted to uncover the inverse mapping in data-driven models to solve inverse problems. A higher degree of sampling over the sensor images (i.e., zero-padding) may further reduce the reconstruction image error and even provide additional advantages, i.e., super-resolution phase retrieval from multiple phase-coded diffraction patterns, [116] and depth detection [117].

Chapter 4

Noise Robust Classification via Multi-Lens Arrays

4.1 Introduction

There has recently been significant interest in the applications of mathematical computation via the propagation of light[118, 62, 63, 64, 65, 66, 67, 68, 17], due to the ever-growing need for "edge computation" and the limits of computational power in a growing number of applications. For example, an autonomous vehicle such as a drone completes image classification using its own CPU instead of sending data to a remote server, and such a scheme causes minimal latency and practical power requirements. Computational schemes that involve optics are considered to be more energy efficient and, in some cases, even "free" in terms of the requisite power consumption. We exploit the fact that the Fraunhofer intensity pattern in the back-focal plane of a lens is the Fourier transform of the

electric field [89]. What we highlight here is the interesting combination of both concepts of edge detection and Fourier optics together, namely, the mathematical momentum-space operations that are useful for machine learning. We explore the limits of performance of the simplest neural nets for classification using spatial phase encoding of a phase vortex.

Vortex light beams with spiral phase gradients are associated with phase singularities, whose phase winding is characterized by a topological charge and singularity around which the electric field is strictly zero [77, 78, 79, 80, 81, 82, 83, 84]. Optical vortices are synonymous with $l \neq 0$, higher-order modes in radially-symmetric cylindrical coordinates. Where the phase is undefined, at the vortex singularity, the electric field is strictly zero. This dark void at the center of optical vortices is exploited in vortex coronography [80] in a manner analogous to the on-axis opaque spot used in dark-field imaging. A recent publication highlights a vortex camera and the the high-resolution that can be achieved with an optical vortex camera [85].

We recently focus on the optical phase vortex for differentiation in compressed, spectral representations [69]. The spectral representations with neural networks are well-known to be efficient with image processing and neural networks [72]. Examples of Fourier-space, deep-learning, diffractive neural networks that are fully analog are demonstrated in [119]. Our technique is a hybrid optical-digital approach similar to that proposed in [65]. The scheme leverages the coherent nature of light for swift image signal processing and is also extremely robust to noise. The dense, shallow, "small-brain" neural networks achieve optimization with low computational cost. A detector that is placed in the Fourier plane of the lens will capture the spectral representation of an image, whose primary spectral

components are concentrated with intensities 200-2000 times higher, which is an effective increase in peak signal to noise (PSNR).

4.2 Results

4.2.1 2D Recursive Relations of the Vortex Fourier Transform

In [69], we demonstrate that the presence of topologically-distinct lenslets provides ptychographic encoding that not only aids solutions to the inverse problem with a neural network, but yields compact representations that are robust to noise. We demonstrate that images are easily reconstructed by shallow, "small-brain"-neural networks when encoded with chiral masks and represented in the Fourier domain. The topological phase is produced by a fixed phase structure on each lens, while each lens produces a distinct Fourier-transform in the lens focal plane. Here, we elaborate on some details that we encountered related to classification of the MNIST handwritten digits.

Firstly, we restate our description of the encoding. A few, small-image datasets are used as object inputs $\mathbf{X}$. These positive, real-valued pixel images are imprinted on an electric field as a phase change:

$$F(r, \phi) = e^{i\alpha_0 \mathbf{X}} \tag{4.1}$$

where α_0 is the dynamic range of the object phase-shift.

The lenslet phase is most easily decomposed as a product of a quadratic radial phase and vortex phase characterized by a topological charge, m,

$$M_m(r, \phi) = \begin{cases} e^{\frac{-i\pi r^2}{\lambda f} + im\phi} & 0 \leq r < a \\ \\ 0 & otherwise \end{cases} \tag{4.2}$$

where r is the radial coordinate, ϕ is the azimuthal coordinate, λ is the wavelength, a is the radius of the phase pattern, and f is the focal length associated with the radial quadratic phase. One of the primary results of our previous paper [69] is that the phase mask $M_m(r, \phi)$ provides edge enhancement for the Gaussian-illuminated object,

$$F'(u, v) \propto \left(\left[\text{sgn}(m)\frac{\partial}{\partial v} - i\frac{\partial}{\partial u} \right] \right)^{|m|} \mathcal{F}\{F(r, \phi)e^{-r^2/w^2}\}, \tag{4.3}$$

where w is the width of a Gaussian beam. This relation is a 2D recursion formula for vortex beams. In Fig. 4.2, we illustrate this result in a simpler form with a Laguerre Gaussian beam p-index $= 0$ beam. The Fourier-plane intensity from the $m = 1$ vortex beam F'_2 is composed of complex-valued spatial derivatives of the output from the Gaussian beam F'_1,

$$F'_2(u, v) \propto \left(\left[\text{sgn}(m)\frac{\partial}{\partial v} - i\frac{\partial}{\partial u} \right] \right)^{|m|} F'_1(u, v). \tag{4.4}$$

A comparison of intensity patterns point to this result. The vortex intensity output pattern [Fig. 4.2 (g)] contains higher intensities where there are zeros in the Gaussian-beam intensity output pattern [Fig. 4.2 (e)].
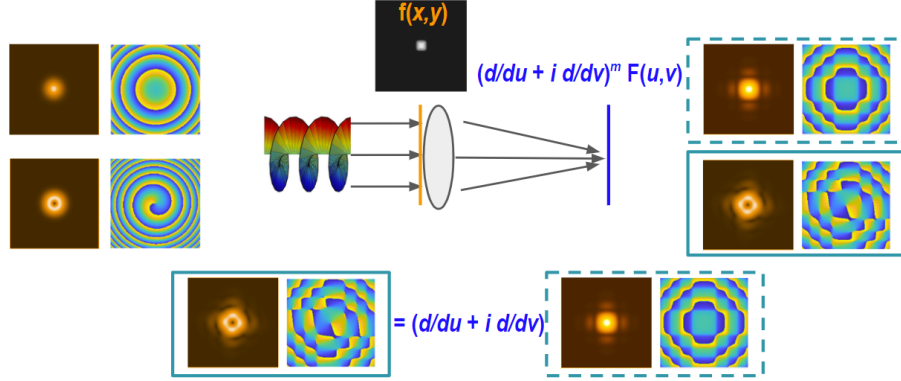
Figure 4.1: Illustration of the edge-enhanced vortex-Fourier pattern in the back-focal plane of the lens. Two input intensity and phase profiles are shown for a Gaussian F_1 and Laguerre Gaussian F_2 beam where the topological charge is m=1. Both beams are focused as a result of the imprinted quadratic radial phase. A sketch demonstrates the input beam illumination of a square obstacle $f(x,y)$ with soft edges. The back focal-plane intensity and phase for Gaussian illumination of $f(x,y)$ is $F(u,v)$, which is related to the back focal-plane intensity for vortex illumination by the relation, $F_{LG} \propto (d/dv - id/du)F$.

In [69], we attribute significant improvements of speed and accuracy of the image reconstruction to the edge-detection scheme. We are able to reconstruct thousands of MNIST images per second on an average 2.0 GHz dual-core CPU with AVX2 instructions. The topologically-distinct phase patterns encode phase information of the Fourier-transformed images into intensity-only patterns. Subsequent post-processing can extract the phase information. While we previously describe image reconstruction with vortex Fourier encoding, we elaborate here on our results with vortex Fourier encoding for MNIST classification problems. Again, with the Fourier representation, we obviate the need for deep-learning algorithms.
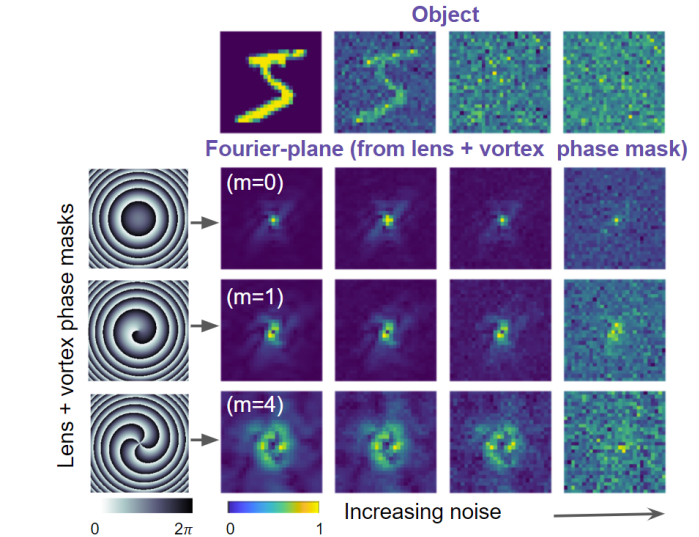
Figure 4.2: The comparison of conventional and vortex-Fourier imaging. The sketch demonstrates that vortex-Fourier map preserves the structural information better than the conventional imaging. The m corresponds to the topological charge of the vortex - the spatial winding speed of the phase of the corresponding beam.

4.2.2 Classification with no Hidden Layers

The MNIST datasets are generally used as a benchmark for categorical classification. With vortex Fourier-transformed images, the most notable improvements are with speed and classification accuracy in the presence of added noise when the neural-network inputs are $\mathbf{Y} = |F'(u, v)|^2$ [Eq. 4.4]. Without noise and when both neural networks are provided with the vortex-encoded Fourier inputs, the small brain is as accurate but learns faster than CNNs. The reason that small-brain dense neural networks are more robust to noise is because they are less computationally complex.

We illustrate this classification robustness with vortex spatial encoding using the MNIST handwritten digit dataset in Fig. 4.3. As before, the inputs $\mathbf{Y}$ are the vortex-Fourier transformed digits, while outputs for classification are the number classes $0 - 9$. No

hidden layers are needed for the classification in order to achieve a similar level of accuracy achieved with convolutional neural networks (CNNs). The accuracies achieved with CNNs compared to multiple-vortex inputs are shown in Fig. 4.3(a).

A confusion plot for a single vortex ($m = 3$) and SNR of -3dB PSNR is shown in Fig. 4.3. The PSNR is calculated at the object, based on [98], which we use in [69]. Different numbers (shapes) attain varying levels of classification output robustness to noise, which is related to our choice of m. In other words, certain digit geometries are more clearly mapped by certain m. We expect that vortex transmutation provides a fingerprint for the neural network classification. Vortex transmutation, the process of vortex charge breakup and migration, is determined by a combination of the object group symmetry and m [120]. Further studies along these lines may identify the weights of the trained small-brain to aid understanding of the learned reconstruction mapping.

In Fig. 4.3 (b) we show the confusion matrix for the case of a three vortex inputs with topological charges $m = 1, 3$, and 5. Again, classification is performed with a single-layer. As a matter of fact, the best performance levels are achieved for the number "1" for all topological charges. An interesting note is that the performance (validation accuracy) on the numbers "6" and "9" differ significantly. This may be explained by the fact that these two numbers tend to be written by hand (and thus represented in the dataset) very differently, despite their "symmetry" in print.

We find that the depth of the phase modulation of the input on the beam α_0 significantly affects the rate of convergence of the classification accuracy [Eq. 4.1]. Using the topological charge $m = 1$, and the scheme with $\alpha_0 = \pi$ achieves 0.95 validation accuracy
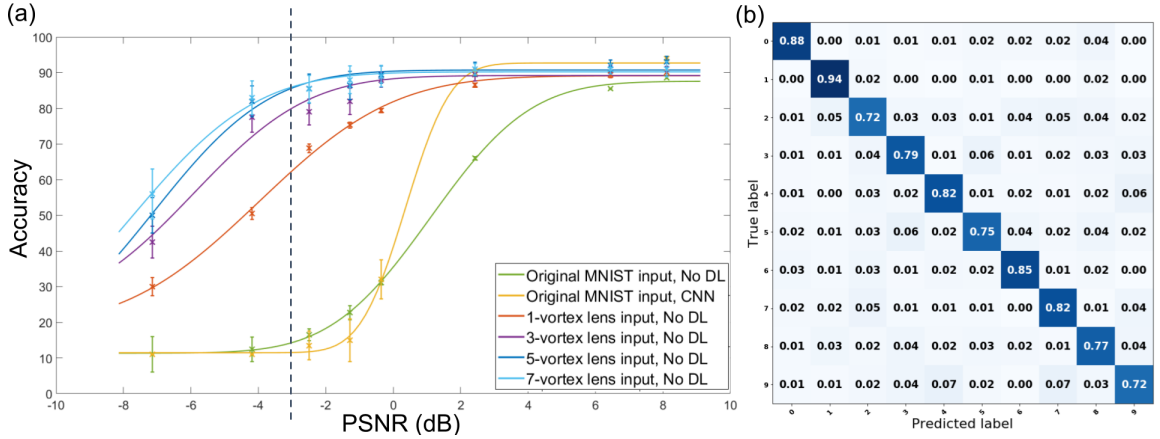
Figure 4.3: (a) Categorical classification accuracy of the MNIST handwritten dataset as a function of PSNR for machine-learning inputs that combine 1, 3, 5, and 7 vortex patterns without deep learning (DL) or with deep convolutional neural networks. When the classification is solved conventionally, with the original MNIST data as inputs to a convolutional neural network (CNN), the categorical accuracy is higher with our scheme when the PSNR is greater than 3 dB. A black dotted line denotes PSNR of -3dB where (b) we plot the corresponding confusion plot for three vortex inputs $m = 1, 3$, and 5 where we achieve 0.8 accuracy.

after 400 training iterations, while the scheme with $\alpha_0 = \pi/2$ achieves the same result after only 240 training iterations. A 0.9 classification accuracy is achieved after the first training iteration, in both cases.

Moreover, we note that there are striking differences in the useful information that the neural network relies on, between these two cases, to classify the digits. With a modulation depth $\alpha_0 = \pi$ is not sensitive to the DC and low-frequency spectral components. In other words, the DC spectral components can be removed from the neural network inputs without affecting the classification accuracy. On the other hand, when the phase modulation depth is $\alpha_0 = \pi/2$, valuable information is concentrated in the DC range. When this DC component is removed from the neural network input, the maximal achievable accuracy

drops to 0.37. This provides valuable insight into the operation of neural nets based on spectral methods.

4.3   Conclusion

The vortex-Fourier encoding scheme provides a critical advantage of compressive sensing, where the number of pixels needed for classification or reconstruction is much lower than that in the case of the conventional imaging. As a result, we are able to achieve classification accuracies >0.80 with a neural network that has no hidden layers with PSNRs as low as -3dB when convolutional neural networks fail. The benefits of the "small-brain" neural network is lower computational complexity, faster computational speeds, and robustness to noise.

In this work we demonstrated the use of shallow neural networks with no hidden layers to classify shapes using extremely weak light intensity sources. The algorithmic approach demonstrated here may be useful for firefighters to see objects through flames [23], to image delicate biological objects prone to photobleaching and phototoxicity[121], healthcare [122], for low-energy surveillance systems [123], energy-efficient traffic control [124], robotics [125], fast velocity object imaging [126] and many other important applications.

Chapter 5

Conclusions

In this work we demonstrated that optical preprocessing can be used to reduce the computational complexity of inverse maps for the Phase Problem. We demonstrate that low SVD-H vortex phase encoders reduce the latencies of reconstruction to sub-millisecond range. Models based on these architectures, in comparison to CNN and U-Net architectures, require three and two orders of magnitude less time for train and test times correspondingly. Although we did not establish the formal procedure for optimization of the preprocessing patterns, we achieved important and useful results.

Our reconstruction and classification schemes excel under conditions of extreme low-light and/or very high levels of noise. They obviously outperform convolutional neural networks in terms of power, latency of reconstruction and noise-robustness. The potential applications are edge computations performed on, for example, various drones. Another role in which our architectures are useful is an "assistant neural network" which performs

low cost image reconstruction and provides a very reasonable initial guess for the more sophisticated and computationally demanding neural network.

We also show that even such simple architectures are capable of generalization, within certain limits. It opens the perspective for majority decisions/reconstructions via "fly eye" conglomerate of reconstructions done with multi-lens systems, which act similarly to ensembles of decision or regression trees (random forests and regression forests). Such systems will be very useful in the high-noise and/or low-light environments, where few failed reconstructions will be "outvoted" by the majority of the correct ones.

Overall, this thesis advances the understanding of the hybrid optical-electronical Machine Vision algorithms for the number of important applications.

References

[1] Tatiana Latychevskaia. Iterative phase retrieval in coherent diffractive imaging: practical issues. *Applied Optics*, 57(25):7187, August 2018.

[2] V. Elser, I. Rankenburg, and P. Thibault. Searching with iterated maps. *Proceedings of the National Academy of Sciences*, 104(2):418–423, January 2007.

[3] Jacopo Bertolotti, Elbert G. van Putten, Christian Blum, Ad Lagendijk, Willem L. Vos, and Allard P. Mosk. Non-invasive imaging through opaque scattering layers. *Nature*, 491(7423):232–234, November 2012.

[4] Ori Katz, Eran Small, Yefeng Guan, and Yaron Silberberg. Noninvasive nonlinear focusing and imaging through strongly scattering turbid layers. *Optica*, 1(3):170, September 2014.

[5] Alex Turpin, Ivan Vishniakou, and Johannes d. Seelig. Light scattering control in transmission and reflection with neural networks. *Optics Express*, 26(23):30911, November 2018.

[6] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3):70, July 2007.

[7] Wanli Chi and Nicholas George. Phase-coded aperture for optical imaging. *Optics Communications*, 282(11):2110–2117, June 2009.

[8] George Barbastathis, Aydogan Ozcan, and Guohai Situ. On the use of deep learning for computational imaging. *Optica*, 6(8):921, July 2019.

[9] Tatsuki Tahara, Xiangyu Quan, Reo Otani, Yasuhiro Takaki, and Osamu Matoba. Digital holography and its multidimensional imaging applications: a review. *Microscopy*, 67(2):55–67, February 2018.

[10] Yair Rivenson, Yichen Wu, and Aydogan Ozcan. Deep learning in holography and coherent imaging. *Light: Science & Applications*, 8(1), September 2019.

[11] Pavan Chandra Konda, Lars Loetgering, Kevin C. Zhou, Shiqi Xu, Andrew R. Harvey, and Roarke Horstmeyer. Fourier ptychography: current applications and future promises. *Optics Express*, 28(7):9603, March 2020.

[12] Pavel Sidorenko and Oren Cohen. Single-shot ptychography. *Optica*, 3(1):9, December 2015.

[13] Fucai Zhang, Isaac Peterson, Joan Vila-Comamala, Ana Diaz, Felisa Berenguer, Richard Bean, Bo Chen, Andreas Menzel, Ian K. Robinson, and John M. Rodenburg. Translation position determination in ptychographic coherent diffraction imaging. *Optics Express*, 21(11):13592, May 2013.

[14] Haiyan Wang and Rafael Piestun. Azimuthal multiplexing 3d diffractive optics. *Scientific Reports*, 10(1), April 2020.

[15] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117, September 2017.

[16] Christopher A. Metzler, Felix Heide, Prasana Rangarajan, Muralidhar Madabhushi Balaji, Aparna Viswanath, Ashok Veeraraghavan, and Richard G. Baraniuk. Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging. *Optica*, 7(1):63, January 2020.

[17] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338–341, March 2018.

[18] Xin Lei, Liangyu He, Yixuan Tan, Ken Xingze Wang, Xinggang Wang, Yihan Du, Shanhui Fan, and Zongfu Yu. 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.

[19] Yichen Wu, Yair Rivenson, Yibo Zhang, Zhensong Wei, Harun Günaydin, Xing Lin, and Aydogan Ozcan. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica*, 5(6):704, May 2018.

[20] Tairan Liu, Kevin de Haan, Yair Rivenson, Zhensong Wei, Xin Zeng, Yibo Zhang, and Aydogan Ozcan. Deep learning-based super-resolution in coherent imaging systems. *Scientific Reports*, 9(1), March 2019.

[21] Yunzhe Li, Yujia Xue, and Lei Tian. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica*, 5(10):1181, September 2018.

[22] Kechen Song, Shaopeng Hu, Xin Wen, and Yunhui Yan. Fast 3d shape measurement using fourier transform profilometry without phase unwrapping. *Optics and Lasers in Engineering*, 84:74–81, September 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. *Optics Express*, 21(5):5379, February 2013.

[24] Eric W. Mitchell, Matthew S. Hoehler, Fabrizio R. Giorgetta, Torrey Hayden, Gregory B. Rieker, Nathan R. Newbury, and Esther Baumann. Coherent laser ranging for precision imaging through flames. *Optica*, 5(8):988, August 2018.

[25] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. DiffuserCam: lensless single-exposure 3d imaging. *Optica*, 5(1):1, December 2017.

[26] Ian Goodfellow, Patrick McDaniel, and Nicolas Papernot. Making machine learning robust against adversarial inputs. *Communications of the ACM*, 61(7):56–66, June 2018.

[27] YongKeun Park, Christian Depeursinge, and Gabriel Popescu. Quantitative phase imaging in biomedicine. *Nature Photonics*, 12(10):578–589, September 2018.

[28] Yujia Xue, Shiyi Cheng, Yunzhe Li, and Lei Tian. Reliable deep-learning-based phase imaging with uncertainty quantification. *Optica*, 6(5):618, May 2019.

[29] R. P. Millane. Phase retrieval in crystallography and optics. *Journal of the Optical Society of America A*, 7(3):394, March 1990.

[30] Ralph W Gerchberg. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[31] James R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

[32] Yoav Shechtman, Yonina C. Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, May 2015.

[33] M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. Flatcam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging*, 3(3):384–397, 2016.

[34] Klaus Wakonig, Ana Diaz, Anne Bonnin, Marco Stampanoni, Anna Bergamaschi, Johannes Ihli, Manuel Guizar-Sicairos, and Andreas Menzel.

X-ray fourier ptychography. *Science Advances*, 5(2):eaav0282, February 2019.

[35] Guoan Zheng, Roarke Horstmeyer, and Changhuei Yang. Wide-field, high-resolution fourier ptychographic microscopy. *Nature Photonics*, 7(9):739–745, July 2013.

[36] Yin Xiao, Lina Zhou, and Wen Chen. Fourier spectrum retrieval in single-pixel imaging. *IEEE Photonics Journal*, 11(2):1–11, April 2019.

[37] Ming-Jie Sun and Jia-Min Zhang. Single-pixel imaging and its application in three-dimensional reconstruction: A brief review. *Sensors*, 19(3):732, February 2019.

[38] Matthew P. Edgar, Graham M. Gibson, and Miles J. Padgett. Principles and prospects for single-pixel imaging. *Nature Photonics*, 13(1):13–20, December 2018.

[39] Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008.

[40] Xinyao Hu, Hao Zhang, Qian Zhao, Panpan Yu, Yinmei Li, and Lei Gong. Single-pixel phase imaging by fourier spectrum sampling. *Applied Physics Letters*, 114(5):051102, February 2019.

[41] Huaxia Deng, Xicheng Gao, Mengchao Ma, Pengcheng Yao, Qingtian Guan, Xiang Zhong, and Jin Zhang. Fourier single-pixel imaging using fewer illumination patterns. *Applied Physics Letters*, 114(22):221906, June 2019.

[42] Zhenyu Liang, Dabin Yu, Zhengdong Cheng, and Xiang Zhai. Adaptive fourier single-pixel imaging sampling based on frequency coefficients prediction. *Optical Engineering*, 59(07):1, July 2020.

[43] Zibang Zhang, Xiao Ma, and Jingang Zhong. Single-pixel imaging by means of fourier spectrum acquisition. *Nature Communications*, 6(1), February 2015.

[44] Matthew H. Seaberg, Alexandre d'Aspremont, and Joshua J. Turner. Coherent diffractive imaging using randomly coded masks. *Applied Physics Letters*, 107(23):231103, December 2015.

[45] Christopher A. Metzler, Philip Schniter, Ashok Veeraraghavan, and Richard G. Baraniuk. prdeep: Robust phase retrieval with a flexible deep network, 2018.

[46] Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, September 2015.

[47] Michael R. Kellman, Emrah Bostan, Nicole A. Repina, and Laura Waller. Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. *IEEE Transactions on Computational Imaging*, 5(3):344–353, September 2019.

[48] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization, 2016.

[49] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, page 201907377, May 2020.

[50] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, May 2017.

[51] Ali Bou Nassif, Ismail Shahin, Imtinan Attili, Mohammad Azzeh, and Khaled Shaalan. Speech recognition using deep neural networks: A systematic review. *IEEE Access*, 7:19143–19165, 2019.

[52] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709, October 2020.

[53] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan. Automatic subspace clustering of high dimensional data. *Data Mining and Knowledge Discovery*, 11(1):5–33, July 2005.

[54] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction, 2018.

[55] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[56] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008.

[57] Sahand Hariri, Matias Carrasco Kind, and Robert J. Brunner. Extended isolation forest. *IEEE Transactions on Knowledge and Data Engineering*, 33(4):1479–1489, April 2021.

[58] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[59] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[60] Luat Vuong and Hobson Lane. Nonlinear spectral preprocessing for small-brain machine learning. In Michael E. Zelinski, Tarek M. Taha, Jonathan Howe, Abdul A. Awwal, and Khan M. Iftekharuddin, editors, *Applications of Machine Learning*. SPIE, September 2019.

[61] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. Learning-based imaging through scattering media. *Optics Express*, 24(13):13738, June 2016.

[62] You Zhou, Hanyu Zheng, Ivan I. Kravchenko, and Jason Valentine. Flat optics for image differentiation. *Nature Photonics*, 14(5):316–323, February 2020.

[63] Erfan Khoram, Ang Chen, Dianjing Liu, Lei Ying, Qiqi Wang, Ming Yuan, and Zongfu Yu. Nanophotonic media for artificial neural inference. *Photonics Research*, 7(8):823, July 2019.

[64] Andrea Cordaro, Hoyeong Kwon, Dimitrios Sounas, A. Femius Koenderink, Andrea Alù, and Albert Polman. High-index dielectric metasurfaces performing mathematical operations. *Nano Letters*, 19(12):8418–8423, November 2019.

[65] Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific Reports*, 8(1), August 2018.

[66] Tengfeng Zhu, Yihan Zhou, Yijie Lou, Hui Ye, Min Qiu, Zhichao Ruan, and Shanhui Fan. Plasmonic computing of spatial differentiation. *Nature Communications*, 8(1), May 2017.

[67] Cheng Guo, Meng Xiao, Momchil Minkov, Yu Shi, and Shanhui Fan. Photonic crystal slab laplace operator for image differentiation. *Optica*, 5(3):251, February 2018.

[68] Xing Lin, Yair Rivenson, Nezih T. Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008, July 2018.

[69] Baurzhan Muminov and Luat 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.

[70] Rosario M Balboa and Norberto M Grzywacz. Power spectra and distribution of contrasts of natural images from different habitats. *Vision Research*, 43(24):2527–2537, November 2003.

[71] Felix Krahmer and Rachel Ward. Stable and robust sampling strategies for compressive imaging. *IEEE Transactions on Image Processing*, 23(2):612–622, February 2014.

[72] Oren Rippel, Jasper Snoek, and Ryan P. Adams. Spectral representations for convolutional neural networks. In Cortes, C and Lawrence, ND and Lee, DD and Sugiyama, M and Garnett, R, editor, *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, volume 28 of *Advances in Neural Information Processing Systems*, 10010 NORTH TORREY PINES RD, LA JOLLA, CALIFORNIA 92037 USA, 2015. Neural Information Processing Systems (NIPS). 29th Annual Conference on Neural Information Processing Systems (NIPS), Montreal, CANADA, DEC 07-12, 2015.

[73] Michele Alberti, Angela Botros, Narayan Schuez, Rolf Ingold, Marcus Liwicki, and Mathias Seuret. Trainable spectrally initializable matrix transformations in convolutional neural networks, 2019.

[74] Călin-Adrian Popa and Cosmin Cernăzanu-Glăvan. Fourier transform-based image classification using complex-valued convolutional neural networks. In *Advances in Neural Networks – ISNN 2018*, pages 300–309. Springer International Publishing, 2018.

[75] Wenlin Chen, James Wilson, Stephen Tyree, Kilian Q. Weinberger, and Yixin Chen. Compressing convolutional neural networks in the frequency domain. In *Proceedings of the 22nd ACM SIGKDD International Confer-

ence on Knowledge Discovery and Data Mining - KDD '16. ACM Press, 2016.

[76] Mathew J. Cherukara, Youssef S. G. Nashed, and Ross J. Harder. Real-time coherent diffraction inversion using deep generative networks. *Scientific Reports*, 8(1), November 2018.

[77] Xuewen Wang, Zhongquan Nie, Yao Liang, Jian Wang, Tao Li, and Baohua Jia. Recent advances on optical vortex generation. *Nanophotonics*, 7(9):1533–1556, August 2018.

[78] 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. *Physical Review A*, 45(11):8185–8189, June 1992.

[79] 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. *Physical Review A*, 56(5):4064–4075, November 1997.

[80] Jae Hoon Lee, Gregory Foo, Eric G. Johnson, and Grover A. Swartzlander. Experimental verification of an optical vortex coronagraph. *Physical Review Letters*, 97(5), August 2006.

[81] Mark R. Dennis, Kevin O'Holleran, and Miles J. Padgett. Chapter 5 singular optics: Optical vortices and polarization singularities. In *Progress in Optics*, pages 293–363. Elsevier, 2009.

[82] Long Zhu and Jian Wang. A review of multiple optical vortices generation: methods and applications. *Frontiers of Optoelectronics*, 12(1):52–68, March 2019.

[83] Shumei Chen, Yuan Cai, Guixin Li, Shuang Zhang, and Kok Wai Cheah. Geometric metasurface fork gratings for vortex-beam generation and manipulation. *Laser & Photonics Reviews*, 10(2):322–326, March 2016.

[84] Gabriel Molina-Terriza, Juan P. Torres, and Lluis Torner. Twisted photons. *Nature Physics*, 3(5):305–310, May 2007.

[85] Kyle Novak and Abbie T. Watnik. Compact vortex wavefront coding camera. In Jonathan C. Petruccelli, Lei Tian, and Chrysanthe Preza, editors, *Computational Imaging V*. SPIE, April 2020.

[86] Xiaoliang He, Cheng Liu, and Jianqiang Zhu. Single-shot fourier ptychography based on diffractive beam splitting. *Optics Letters*, 43(2):214, January 2018.

[87] William N. Plick and Mario Krenn. Physical meaning of the radial index of laguerre-gauss beams. *Physical Review A*, 92(6), December 2015.

[88] Saifollah Rasouli and Davud Hebri. 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(5):800, April 2019.

[89] J.W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Series in Electrical and Computer Engineering: Communications and Signal Processing. McGraw-Hill, 1996.

[90] Manuel Guizar-Sicairos and James R. Fienup. Holography with extended reference by autocorrelation linear differential operation. *Optics Express*, 15(26):17592, 2007.

[91] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, 2017.

[92] Tarin Clanuwat, Mikel Bober-Irizar, Asanobu Kitamoto, Alex Lamb, Kazuaki Yamamoto, and David Ha. Deep learning for classical japanese literature, 2018.

[93] Hassan M. Najadat, Ahmad A. Alshboul, and Abdullah F. Alabed. Arabic handwritten characters recognition using convolutional neural network. In *2019 10th International Conference on Information and Communication Systems (ICICS)*. IEEE, June 2019.

[94] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. http://yann.lecun.com/exdb/mnist/, 2010.

[95] Nicholas Baker, Hongjing Lu, Gennady Erlikhman, and Philip J. Kellman. Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, 14(12):e1006613, December 2018.

[96] Zhou Wang and A.C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, January 2009.

[97] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, pages 234–241. Springer International Publishing, 2015.

[98] Katsushi Ikeuchi, editor. *Computer Vision: A Reference Guide*. Springer Reference. Springer, New York, 2014.

[99] Changyan Zhu, Eng Aik Chan, You Wang, Weina Peng, Ruixiang Guo, Baile Zhang, Cesare Soci, and Yidong Chong. Image reconstruction through a multimode fiber with a simple neural network architecture, 2020.

[100] Adam Greengard, Yoav Y. Schechner, and Rafael Piestun. Depth from diffracted rotation. *Optics Letters*, 31(2):181, January 2006.

[101] Sri Rama Prasanna Pavani and Rafael Piestun. High-efficiency rotating point spread functions. *Optics Express*, 16(5):3484, February 2008.

[102] Siddharth Malik, Anjali Sardana, and Jaya. A keyless approach to image encryption. In *2012 International Conference on Communication Systems and Network Technologies*. IEEE, May 2012.

[103] Christopher A. Metzler, Felix Heide, Prasana Rangarajan, Muralidhar Madabhushi Balaji, Aparna Viswanath, Ashok Veeraraghavan, and Richard G. Baraniuk. Deep-inverse correlography: towards real-time high-resolution non-line-of-sight imaging. *Optica*, 7(1):63–71, Jan 2020.

[104] Wangyu Luo, Wael Alghamdi, and Yue M Lu. Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Transactions on Signal Processing*, 67(9):2347–2356, 2019.

[105] Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015.

[106] Yi Luo, Deniz Mengu, Nezih T. Yardimci, Yair Rivenson, Muhammed Veli, Mona Jarrahi, and Aydogan Ozcan. Design of task-specific optical systems using broadband diffractive neural networks. *Light: Science & Applications*, 8(1), December 2019.

[107] Yichen Shen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljačić. Deep learning with coherent nanophotonic circuits. *Nature Photonics*, 11(7):441–446, June 2017.

[108] Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David A. B. Miller, and Demetri Psaltis. Inference in artificial intelligence with deep optics and photonics. *Nature*, 588(7836):39–47, December 2020.

[109] Demetri Psaltis and Nabil Farhat. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Optics Letters*, 10(2):98, February 1985.

[110] S. Jutamulia and F.T.S. Yu. Overview of hybrid optical neural networks. *Optics & Laser Technology*, 28(2):59–72, March 1996.

[111] Jiachen Wu, Hua Zhang, Wenhui Zhang, Guofan Jin, Liangcai Cao, and George Barbastathis. Single-shot lensless imaging with fresnel zone aperture and incoherent illumination. *Light: Science & Applications*, 9(1), April 2020.

[112] Shuiqin Zheng, Xuanke Zeng, Lang Zha, Huancheng Shangguan, Shixiang Xu, and Dianyuan Fan. Orthogonality of diffractive deep neural networks, 2018.

[113] Q. R. Razlighi and N. Kehtarnavaz. A comparison study of image spatial entropy. In Majid Rabbani and Robert L. Stevenson, editors, *Visual Communications and Image Processing 2009*. SPIE, January 2009.

[114] Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4):413–424, July 1986.

[115] Mo Deng, Shuai Li, Zhengyun Zhang, Iksung Kang, Nicholas X. Fang, and George Barbastathis. On the interplay between physical and content priors in deep learning for computational imaging. *Optics Express*, 28(16):24152, July 2020.

[116] Vladimir Katkovnik, Igor Shevkunov, Nikolay V. Petrov, and Karen Egiazarian. Computational super-resolution phase retrieval from multiple phase-coded diffraction patterns: simulation study and experiments. *Optica*, 4(7):786, July 2017.

[117] Yi Hua, Shigeki Nakamura, M. Salman Asif, and Aswin C. Sankaranarayanan. SweepCam — depth-aware lensless imaging using programmable masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(7):1606–1617, July 2020.

[118] Demetri Psaltis and David Casasent. Deformation invariant optical processors using coordinate transformations. *Applied Optics*, 16(8):2288, August 1977.

[119] Tao Yan, Jiamin Wu, Tiankuang Zhou, Hao Xie, Feng Xu, Jingtao Fan, Lu Fang, Xing Lin, and Qionghai Dai. Fourier-space diffractive deep neural network. *Physical Review Letters*, 123(2), July 2019.

[120] Albert Ferrando, Mario Zacarés, Miguel-Ángel García-March, Juan A. Monsoriu, and Pedro Fernández de Córdoba. Vortex transmutation. *Physical Review Letters*, 95(12), September 2005.

[121] Chawin Ounkomol, Sharmishtaa Seshamani, Mary M. Maleckar, Forrest Collman, and Gregory R. Johnson. Label-free prediction of three-dimensional fluorescence images from transmitted light microscopy. March 2018.

[122] Sebastian T. Schindera, Devang Odedra, Syed Arsalan Raza, Tae Kyoung Kim, Hyun-Jung Jang, Zsolt Szucs-Farkas, and Patrik Rogalla. Iterative reconstruction algorithm for CT: Can radiation dose be decreased while low-contrast detectability is preserved? *Radiology*, 269(2):511–518, November 2013.

[123] Tian He, Bruce Krogh, Sudha Krishnamurthy, John A. Stankovic, Tarek Abdelzaher, Liqian Luo, Radu Stoleru, Ting Yan, Lin Gu, and Jonathan Hui. Energy-efficient surveillance system using wireless sensor networks. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services - MobiSYS '04*. ACM Press, 2004.

[124] Marcin Lewandowski, Bartłomiej Płaczek, Marcin Bernas, and Piotr Szymała. Road traffic monitoring system based on mobile devices and bluetooth low energy beacons. *Wireless Communications and Mobile Computing*, 2018:1–12, July 2018.

[125] Hatem Alismail, Michael Kaess, Brett Browning, and Simon Lucey. Direct visual odometry in low light using binary descriptors. *IEEE Robotics and Automation Letters*, 2(2):444–451, April 2017.

[126] Paramanand Chandramouli, Samuel Burri, Claudio Bruschini, Edoardo Charbon, and Andreas Kolb. A bit too much? high speed imaging from sparse photon counts, 2018.