Computational Imaging and Analysis in Breast Cancer

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAR 06 2018
LIBRARIES
ARCHIVES

by

Justin Lee

Submitted to the Department of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Health Sciences and Technology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

© Massachusetts Institute of Technology 2018. All rights reserved.

Signature redacted

Author .....
Department of Health Sciences and Technology
Feb 1, 2018

Signature redacted

Certified by .....
George Barbastathis
Professor of Mechanical Engineering, MIT
Thesis Supervisor

Signature redacted

Accepted by .....
.....
Emery N. Brown
Professor of Health Sciences and Technology and
Professor of Computational Neuroscience, MIT
Director, Harvard-MIT Program in Health Sciences and Technology

MIT Libraries

77 Massachusetts Avenue
Cambridge, MA 02139
http://libraries.mit.edu/ask

DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

The images contained in this document are of the best quality available.

Computational Imaging and Analysis in Breast Cancer

by

Justin Lee

Submitted to the Department of Health Sciences and Technology
on Feb 1, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Health Sciences and Technology

Abstract

The conventional pathologic analysis of malignancies involves a qualitative characterization and integration of several factors including tumor size, general degree of differentiation, tumor heterogeneity, mitotic rate, and lymphovascular invasion. For some cancers, biomarkers such as hormone receptor expression or receptor kinase over-expression can provide additional prognostic and therapeutic guidance. Unfortunately, all of these qualitative histologic approaches, while generally accepted for directing patient care, often exhibit significant inter-observer variability resulting in inconsistent inter- and intra-institutional predictions of tumor behavior (including metastases and/or recurrence), resulting in incorrect diagnoses or treatment.

Because cellular morphology is an integrated reflection of genetic and epigenetic expression, we hypothesize that a more accurate quantitative accounting and measurement of histologic features can provide a more robust and reliable prediction of tumor behavior. Computational imaging utilizes software to augment or replace the role of traditional optical elements in imaging systems and has an ability to significantly increase the accuracy, robustness and cost-efficiency of digital pathology. In this thesis, we develop and test three novel computational imaging algorithms including, to the best of our knowledge, the first system for lensless computational imaging through deep learning. We then test our hypothesis by applying augmented image retrieval, analysis algorithms, and machine learning on a validated dataset of breast cancer images where the clinical outcomes of the primary tumor are known. In particular, we analyze algorithms related to identifying mitoses as a central proof of concept.

Thesis Supervisor: George Barbastathis
Title: Professor of Mechanical Engineering, MIT

Acknowledgments

I dedicate this work to my loving wife, Diana, and my son, Liam.

This work would not have been possible without the help and support of a multitude of people.

First and foremost, I would like to thank my research advisor, Professor George Barbastathis for high incredible kindness, mentorship, and patience with me while I meandered through graduate studies.

I would like to thank my committee members, Professor Richard Mitchell and Professor Timothy Padera for taking an interest in my research and for serving on my committee. I have greatly appreciated your clinical insights and advice.

I would like to thank Irina Gaziyeva for the amazing work she has done behind the scenes to allow our group to focus more on research and less on paperwork.

I am grateful for all of my past and present labmates in the 3D optical systems group: Yuan Luo for teaching me the basics of experimental laboratory optics. Laura Waller for introducing me to OSA "optics fun day" and mentoring me early on. Yi Liu, Nader Shaar, Nick Loomis, Jose A. Dominguez-Caballero, Se Baek Oh, Hyungryul Choi, Hanhong Gao, Jonathan Petruccelli, Max Hsieh, and Jeong-gil Kim for helpful office discussions (sometimes on labwork and many other times on totally unrelated matters).

I would like to thank Zhengyun Zhang for his helpful comments and insights at group meeting and Ayan Sinha for patiently explaining questions I had on deep learning.

I would like to give an extra special thanks to Lei Tian and Shuai Li for their help with research.

I am grateful to my former HST classmates: Adam Pan, Kelli Xu, Nikhil Vadhavkar, Vyas Ramanan, and Andrew Warren for being amazing friends, even when I am horrible at keeping in contact with people.

Outside of lab, I would like to thank Bill Herrington for optics teaching demo discussions.

I am grateful for the Computational Science Graduate Fellowship which has sponsored my intellectual curiosity over the course of my studies.

Finally, I would like to thank the community at MIT. I have been a student at MIT for over a third of my life and consider it my second home. Although I'm eager to begin the next chapter of my life, I will always look back upon my tenure at MIT with the fondest of memories.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The conventional pathologic analysis of malignancies involves a qualitative characterization and integration of several factors including tumor size, general degree of differentiation, tumor heterogeneity, mitotic rate, and lymphovascular invasion. For some cancers, biomarkers such as hormone receptor expression or receptor kinase over-expression can provide additional prognostic and therapeutic guidance. Unfortunately, all of these qualitative histologic approaches, while generally accepted for directing patient care, often exhibit significant inter-observer variability resulting in inconsistent inter- and intra-institutional predictions of tumor behavior (including metastases and/or recurrence), resulting in incorrect diagnoses/treatment [1, 2].

Because cellular morphology is an integrated reflection of genetic and epigenetic expression, we hypothesize that a more accurate quantitative accounting and measurement of histologic features can provide a more robust and reliable prediction of tumor behavior. We propose to specifically test this hypothesis by applying augmented image retrieval, analysis algorithms, and machine learning on a validated dataset of breast cancer images where the clinical outcomes of the primary tumor are known. In particular, we will analyze algorithms related to identifying mitoses as a central proof of concept.

This thesis explores algorithms and applications of "computational imaging" to pathology, which can enable pathologists to image faster, cheaper, more robustly, and also provide augmented information about the sample being imaged.

Chapter Two provides a brief background on clinical motivations for the work in this thesis and reviews the current state of imaging in pathology.

Chapter Three provides background theory on conventional microscopic imaging and computational microscopy techniques. It also covers background theory on noisy signal recovery, regularization, and deep learning.

Chapter Four describes phase-space analysis of light fields and presents two novel applications of algorithms for low-rank matrix recovery to improve signal recovery in two phase-space imaging techniques: phase-space tomography (PST) and Wigner distribution deconvolution (WDD).

Chapter Five reviews three lensless imaging techniques and presents simulation and experimental results of, to the best of our knowledge, the first system ever for lensless computational imaging through deep learning.

Chapter Six reviews ptychography (and its Fourier dual), presents an experimental setup for automated mitoses detection in a desktop microscope, and details an experiment testing the hypothesis that deep learning from a focal stack of images (in-focus image + "lateral flux" image) can provide increased mitosis detection accuracy.

Chapter 2

Clinical Motivations

2.1 Breast cancer and its clinical management

Breast cancer is the most frequency malignancy in women in both the developed and developing world. In the United States, the incidence of breast cancer is 118.7/100,000 women per year [3], and 1 in 8 U.S. females will develop breast cancer in her lifetime [4]. Although earlier detection through mammographic screening, increased awareness, and advances in effective treatments have improved the outcome for many women, breast cancer remains a significant burden on society with the disease claiming over 40,000 U.S. lives annually [4]. Worldwide, breast cancer affects about 12% of all women and in 2012, over 1.6 million new cases were diagnosed [5].

2.1.1 Risk Factors

Lifestyle risk factors for breast cancer include: smoking tobacco [6], poor diet [7], use of hormonal birth control [8, 9], exposure to radiation, and shift work [10]. In a small minority of breast cancer cases (5-10%), genetics (e.g., BRCA1/2 mutation) is believed to be the primary cause [11].

2.1.2 Screening

Breast cancer is commonly diagnosed through screening or a noticed symptom (e.g., pain or palpable mass in the breast). Mammography is a common screening method for breast cancer and is recommended by the American Cancer Society [12] at age 45 or sooner. Positive impacts of screening mammography include decreased breast cancer mortality (15% for women in their 40s and 32% for women in their 60s). Negative impacts include: radiation exposure, false-positive examinations, and anxiety. The risk of a false-positive mammogrography over a 10-year period for women screened in their 40s is 61%, and this false-positive risk decreases with age [12].

Collectively, physical examination, mammography, and microscopic evaluation are used to diagnose breast cancer. When physical examination and mammography are inconclusive, a sample in the lump may be acquired via fine needle aspiration (FNA) and examined under brightfield microscopy to help establish a diagnosis. For instance, clear fluid is suggestive that the lump is likely not cancerous, while microscopic observation of cancerous cells suggests a more serious diagnosis. In addition to FNA, microscopy slides prepared from core biopsy or excisional biopsy samples may also be examined under brightfield microscopy.

2.1.3 Classification and Treatment

Breast Cancers are commonly classified by:

- **Stage**

  Breast cancer staging is commonly done using the TNM system, which stages breast cancer based on tumor size (T), presence/absence of axillary lymph node metastases (N), and whether or not the tumor has metastasized (M). Larger size, nodal spread and presence of metastases indicate a worse prognosis.

- **Histopathology** (histologic appearance)

  The three most common histopathological breast cancer types are [13]:

  1. Invasive ductal carcinoma (55% of U.S. breast cancers)

2. Ductal carcinoma in situ (13% of U.S. breast cancers), and

3. Invasive lobular carcinoma (5% of U.S. breast cancers)

- **Receptor status**

  ER, PR, and HER2 receptor status are routinely assessed (primarily by immunohistochemistry), and targeted therapies exist for patients with e.g., HER2+ status (trastuzumab, an anti-HER2 antibody) or ER+ status (anti-estrogen hormonal therapy). Example guidelines for treatment based on breast cancer receptor status and grade are shown in the table below [14]:

1. **[ER+/HER2+]** "Luminal B/C" (15-20% of No-Special-Type/ductal carcinoma)
   Grade 1 -------> Possible hormone therapy
   Grade 2 -------> Hormone therapy; possible chemotherapy + anti-HER2 Ab (e.g., trastuzumab)
   Grade 3 -------> Anti-HER2 Ab + hormone therapy + chemotherapy
2. **[ER+/HER2-]** "Luminal A" (40-55% of NST) + "Normal Breast-like" (6-10% of NST)
   Grade 1 -------> Possible hormone therapy
   Grade 2 -------> Hormone therapy + possible chemotherapy
   Grade 3 -------> Hormone therapy and chemotherapy
3. **[ER-/HER2+]** "HER2 positive", (7-12% of NST cancers)
   Grade 1 -------> Possible Anti-HER2 Ab (e.g., trastuzumab)
   Grade 2 -------> Anti-HER2 Ab + possible chemotherapy
   Grade 3 -------> Anti-HER2 Ab + chemotherapy
4. **[ER-/HER2-]** "Basal-like" (13-25% of NST cancers)
   Grade 1 -------> No further treatment (if no cancer in lymph nodes), possible chemotherapy
   Grade 2 -------> Possible chemotherapy
   Grade 3 -------> Chemotherapy

- **Grade**

  When grading a cancer, pathologists qualitatively assess (under brightfield microscopy) how the overall tissue architecture looks (e.g., presence/absence of differentiation/gland formation) and how "ugly" individual cells look (e.g., presence of pleomorphisms). These qualitative assessments are accompanied by a (semi-) quantitative measure of mitotic counts. Mitotic figures are counted manually by pathologists by viewing several high-magnification fields of view under brightfield microscopy. Quantificiation of the proliferative activity of tumors is of interest to pathologists because the proliferation rate of a tumor has been shown to strongly correlate with degree of tumor aggressiveness and to be prognostic of clinical outcome [15, 16].

Genetic testing of tumor tissue samples (e.g., OncotypeDX, Mammoprint, etc.) has been successfully commercialized as a means to predict patient risk for relapse and to help decide if chemotherapy is necessary, but clinical utility of these tests have been limited by their high cost compared with routine IHC and their lack of availability in many environments [15]. Additionally, their independent prognostic utility is still unclear [17] as the majority of markers in commercially available molecular/genomic tests are proliferation-driven.

Standard management of breast cancer involves surgical removal of affected areas of the breast (e.g., lumpectomy or mastectomy) and excision of lymph nodes to check for metastasis. Depending on the cancer's classification (stage, histologic appearance, receptor status, grade), post-surgical adjuvant therapy (chemotherapy, hormonal therapy, targeted therapy, or radiation therapy) may be indicated. Additionally, some breast cancer patients may receive neoadjuvant (pre-operative) chemotherapy.

Optimal patient care involves both avoidance of overtreatment (e.g., the patient's tumor would have been cured solely with surgical excision and hormonal therapy, but a systemic chemotherapy was prescribed) as well as avoidance of undertreatment (e.g., the patient is not given adjuvant therapy and the cancer recurs or the patient is treated with drugs that are ultimately ineffective). The kind of treatment a patient receives matters as well (e.g., patients with HER2-positive tumors should receive targeted anti-HER2 treatments and chemotherapy). Any information that helps clinicians better determine what degree and type of treatment a patient should receive will enable physician to better tailor treatments for individual breast cancer patients.

2.2 Whole-slide imaging

Brightfield microscopy is the current gold-standard for pathological examination of tissue sections or smears. Automated whole-slide imaging (WSI) involves robotically scanning and digitizing an entire histology slide under brightfield microscopy. After image acquisition, image-stitching algorithms are applied to merge each field of view

captured into a digitized "whole-slide." The first automated, high-resolution WSI system was developed by Wetzel and Gilbertson in 1999 [18]. This system was built upon a traditional light microscopy setup ("shift-and-stitch" whole slide imaging) and had a primary magnification of 20x, a numerical aperture of 0.7, and a detector pixel size of 6.6um [18]. In the past two decades, numerous companies have improved upon this system, and many modern WSI scanners utilizing linescan camera synced via time-delay integration (TDI) can image entire slides at 100x within a few minutes).

Significant advantages of WSI over conventional microscopy include: (1) accessibility (images can be viewed anywhere at any time), (2) ease of sharing and retrieval of archival images, and (3) ability to utilize computer-aided diagnostic tools [18, 19, 20, 21]. WSI has been successfully used for: educational purposes (e.g., in digital slide teaching sets), quality assurance (e.g., archiving), and research. However, there is currently no whole slide imager that is FDA-approved for determining primary clinical diagnosis [20, 21] (i.e., at this point in time, WSI may only be used for second opinion consultations). Additionally, due to its high cost (a single WSI imaging system can run over $500,000), it is likely that in the near future WSI will remain constrained to advanced clinical settings. Nevertheless, the field of pathology is trending towards digital integration, and WSI will undoubtedly play a significant role in the future of clinical pathology [22].

While not an explicit focus of this thesis, it should be noted that algorithms and techniques developed in this work may be used to augment current whole-slide imaging processes or even completely reimagine whole slide imaging from a "lenless computational imaging" perspective. Such computational whole-slide imaging approaches may help overcome limitations of traditional brightfield microscopy WSI systems including: relatively slow imaging speed due to limited field-of-view (trading off with spatial resolution), cost, and an ability to observe only amplitude information of the imaged sample.

Chapter 3

Background Theory

3.1 Classical Microscopy

3.1.1 Brightfield and Phase Contrast Microscopy

Brightfield imaging is the simplest form of optical microscopy. In a brightfield microscope setup, a sample is illuminated by a light source and then imaged with magnification onto a camera sensor. A standard brightfield microscope can be thought of as a $4f$ system with object specimen placed at the input plane, a microscope objective lens as $L1$, a second lens further down the optical train as $L2$, and no special filters at the pupil plane.

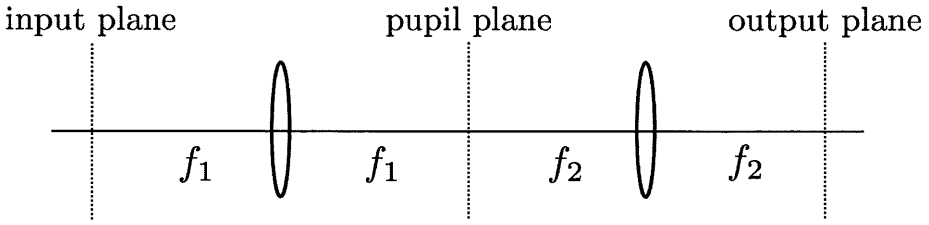

Figure 3-1: $4f$ system.

Thus, a thin specimen or object $o(x, y)$ placed at the input (object) plane of a brightfield microscope will be imaged onto the output plane where a time-averaged intensity $I(x', y')$ can be observed or recorded by a camera detector. Notably, this

output image will contain no phase information from the input object (i.e., phase-shifts induced by the specimen are invisible to the detector).

Many techniques have be used to recover this missing phase information. The most famous of these techniques is phase contrast microscopy, invented by Frits Zernike in the 1930s (for which he would later win a Nobel Prize) [23]. In phase contrast microscopy, a pupil mask "bump" is placed at the Fourier plane of the optical system; this mask mixes previously invisible/undetectable phase-information into observable amplitude variations in the output image. These variations allow observation of phase shifts induced by the sample but are not quantitative in nature.

3.1.2 Numerical Aperture, Magnification, Space Bandwidth Product, Depth of Field

The Numerical Aperture (NA) of an imaging system is given as $n \sin\theta$, where n is the index of refraction of the surrounding medium and θ is the maximum angle of light that the system can capture. This "maximum angle of light" captured by the imaging system determines its lateral resolution (one needs higher angles of light in order to reconstruct higher spatial frequencies of an object), and lateral resolving power is often described in terms of a system's numerical aperture. Additionally, the two-point resolution distance using the Rayleigh criterion is inversely proportional to the NA of an imaging system $\left(R \sim \frac{\lambda}{\mathrm{NA}}\right)$.

Microscope magnification describes the apparent size-increase of an imaged object compared with its actual size. Magnification is related to but not directly correlated with NA/optical resolution (e.g., although the most common 20X magnification microscope objectives have an NA of 0.5, there are exist some 20X objectives with NA=0.75 and many 40X objectives only have an NA=0.75).

The Field of View (FOV) of an imaging system describes the spatial extent of an object that can be seen and is calculated for a microscope as:

$$\mathrm{FOV} = \frac{\text{Field number (typically } \sim 25\text{mm)}}{\text{Magnification}}$$

FOV increases when magnification is decreased and vice versa. As mentioned above, a lower objective magnification does not necessarily correspond to lower imaging resolution and it is possible to have a system simultaneous (relatively) high FOV and resolution.

"Space-Bandwidth Product" (SBP) describes the number of pixels required to capture the full FOV at full resolution. For example, a 20X/0.75NA microscope with a 25mm field number has a FOV of 1.25mm in diameter, a two-point resolution limit of $0.61\frac{\lambda}{\text{NA}} \approx 432\text{nm}$ for blue light ($\sim$ 532nm). To Nyquist-sample this spatial frequency, the (demagnified camera) pixel dimension must be $\leq$ 216nm. Thus, the number of pixels required to capture the full diameter of the FOV at full resolution is: $\frac{1.25\text{mm}}{216\text{nm}} \approx 5787$, and the total number of pixels required to capture a circular field of view would be: $\approx$ 26 million.

3.2 Computational Imaging

Traditional brightfield imaging techniques have several limitations. The first is physics, specifically the diffraction limit, which inherently limits the optical resolution of brightfield microscopy, such that even with an infinitesimally small detector pixel size, one cannot resolve any features smaller than the diffraction limit of the microscope objective. Second, although high NA objectives exist, these objectives are typically expensive, limit the viewer to a very tiny field-of-view, and also have an extremely small depth-of-field, requiring high-precision mechanical movements and calibration in order to acquire quality images. Finally, as mentioned earlier, these conventional imaging systems also cannot retrieve any information about the phase of a light field.

In a conventional imaging setup, the objective of the system is simply to make an exact replica (perhaps magnified) of the object onto the output/camera plane. In contrast, in a computational imaging system, the goal isn't to try to make an exact copy of the object at the sensor plane, but instead to pass as much information as you can through the system given application constraints (e.g., imaging speed, system

cost, robustness, and/or resolution requirements). In a computational imaging system any component can be changed or removed; illumination can be coded, abiitrary filters can be placed at the pupil plane, and there are even computational imaging setups (including several discussed in this work) where all lenses in the "imaging" system are removed altogether.

Computational imaging utilizes software to augment or replace the role of traditional optical elements in imaging systems. By designing imaging hardware and software in concert, imaging systems can be made simpler/more robust and have capabilities beyond that of traditional imaging systems. Examples of popular computational imaging setups include: digital holography (Chapter 4), transport of intensity imaging (Chapter 4), and iterative phase retrieval techniques such as the Gerchberg-Saxton Algorithm (Chapter 4), and Ptychography (Chapter 5) :

3.3 Regularization and Sparsity Priors

Many computational imaging inverse problems are ill-posed. In order to solve these problems, regularizers are often employed. With regularization, an ill-posed problem can be cast as [24]:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ \|Ax - b\|^2 + \alpha\Phi(x) \right\}$$

where the first term $\|Ax-b\|^2$ is a "fitness term" representing how well the estimate for x matches the observed data y and the second term $\alpha\Phi(x)$ is a weighted regularizer. α may be tuned to apply more or less regularization.

Two regularization techniques employed in this thesis include compressive sensing and low-rank matrix recovery.

3.3.1 Compressive Sensing

Compressed Sensing (aka Compressive Sensing) is a method for solving underdetermined linear systems given a strong prior of input "sparsity" in some domain [24, 25, 26]. Specifically, given the linear system $Ax = b$ where A is rank-deficient

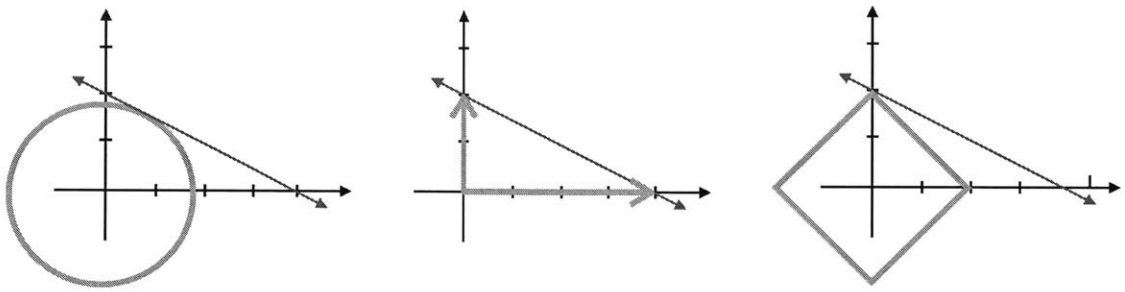

Figure 3-2: Compressive Sensing Reconstruction based on knowledge that an underlying signal is "sparse". (left) L_2 minimization "grows a circle" until it touches the plane of possible solutions, yielding an incorrect answer. (middle) L_0 minimization "searches along the axes" and would yield the correct solution but is NP-hard. (right) L_1 minimization "grows a diamond" and reaches the same solution as the L_0 norm with overwhelming probability in cases where the measurements are incoherent with each other and the restricted isometry property is satisfied.

(e.g., $m < n$), any solution (if one exists) will not unique but instead restricted to lie on a line/plane/hyperplane representing the nullspace of A. The classical solution to estimating a solution to this problem is to minimize the L2 norm of x (i.e., minimize the "energy" of x). This is equivalent to growing a circle/sphere/hypersphere until it touches the line/plane/hyperplane of possible solutions. The first point of contact is the solution that minimizes the L2 norm (while matching the provided constraints).

In contrast, compressive sensing (CS) theory posits that instead of minimizing the L2 norm, one should instead minimize the L0 norm of the signal in order to find the "sparsest" solution that matches the given constraints. In practice, since L0-norm minimization is NP-hard, a convex relaxation is applied and CS algorithms minimize L1-norms. [24, 25, 26]. The CS "object sparsity" assumption is widely applicable since almost all objects (anything that's not pure noise) can be represented "sparsely" in some basis (e.g., wavelets). If the class of objects is known (e.g., you know the images will all be of blood vasculature), then a specialized dictionary or mixed basis can be trained and utilized for image reconstruction/denoising based on compressive sensing priors. In this manner, one can seemingly break the Shannon-Nyquist limit to resolution (because of the a priori knowledge that the object is sparse in a particular

basis).

In practice, minimization of an object's total variation (TV) norm (L1 norm of the object gradient) is commonly used in lieu of tailored bases as TV-norm minimization provides suitable regularization under the assumption that a given object is "smooth".

3.3.2 Low-rank matrix recovery

Low-rank matrix recovery has similar roots to compressive sensing. In the "noisy matrix completion problem," one tries to recover an underlying low-rank matrix M from noisy and possibly incomplete measurements of its entries [27]. M can by any real or complex matrix with dimensions $m \times n$ and rank $r \ll m, n$. Additive or applied noise can be modeled as perturbations to each matrix entry such that:

$$\tilde{M}_{ij} = M_{ij} + Z_{ij} \tag{3.1}$$

where the matrix Z accounts for the added/applied noise and the matrix $\tilde{M}$ is the noisy "approximately low-rank" matrix that is (partially) observed or measured. Of the complete matrix $\tilde{M}$, only a subset Ω of its entries may actually be observed. Noisy matrix completion algorithms seek to find a low-rank estimation X of the original matrix M based on the observed indices Ω of the noisy matrix $\tilde{M}$.

In the case where there is no noise ($\tilde{M} = M$), the following optimization may be used to recover the low-rank matrix exactly, provided that $|\Omega|$ is large enough and satisfies certain "incoherence" conditions [24, 27]:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \text{rank}(X) \\ \text{subject to} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \end{aligned} \tag{3.2}$$

where $\mathcal{P}_\Omega(\cdot)$ is the operator selecting which entries of the matrix we have measured and is defined as:

$$\mathcal{P}_\Omega(M)_{ij} = \begin{cases} M_{ij}, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

Since the problem in (3.2) is NP-hard, a common approach is to apply a convex

relaxation [24, 27, 72] which turns the problem into:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \|X\|_* \\ \text{subject to} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \end{aligned} \tag{3.4}$$

where $\|X\|_*$ denotes the nuclear norm of X (i.e., the sum of all the singular values of X).

Adding noise (which may be adversarial/worst-case), the problem in (4.11) becomes [27, 74, 75]:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \|X\|_* \\ \text{subject to} \quad & \|\mathcal{P}_\Omega(X - \tilde{M})\|_F \leq \delta \end{aligned} \tag{3.5}$$

which can be alternatively expressed in Lagrangian form as:

$$\text{minimize} \quad \mu\|X\|_* + \frac{1}{2}\|\mathcal{P}_\Omega(X - \tilde{M})\|_F^2 \tag{3.6}$$

If certain properties are known about the noise, then we may change the constraints in the above equations to reflect our apriori information. For example, if the noise is impulsive (e.g., dead pixels on a camera), then one may use the entrywise L_1 norm in place of the Frobenius norm in the equations above ($\|\mathcal{P}_\Omega(X - Y)\|_F \rightarrow \|\mathcal{P}_\Omega(X - Y)\|_{L1}$). Likewise, if the noise is known to be low-rank, then we may use the nuclear norm in place of the Frobenius norm.

3.4 Deep Learning

Artificial Neural Networks (ANNs) are a class of machine learning algorithms that utilize artificial "neurons" connected to each other. In a typical ANN, each neuron first multiplies inputs by pre-established weights, then sums these weighted inputs (plus a bias term) and passes the sum through a non-linear "activation" function. The output of the neuron is then passed along to downstream neurons. [28, 29]

Neural networks are "trained" by providing them with example inputs and their

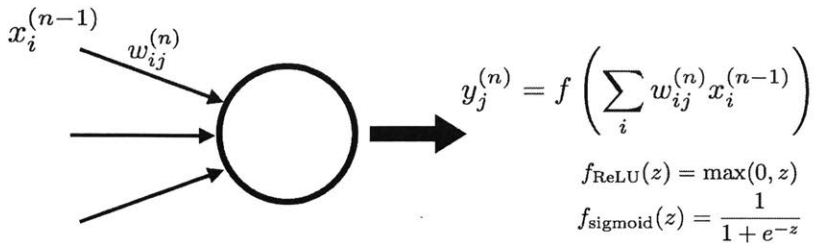

Figure 3-3: Each neuron in an ANN multiplies its inputs by a pre-established weights, then sums the weighted inputs together and passes the sum through a nonlinear activation function. Common activation function include the sigmoid function and rectified linear unit.

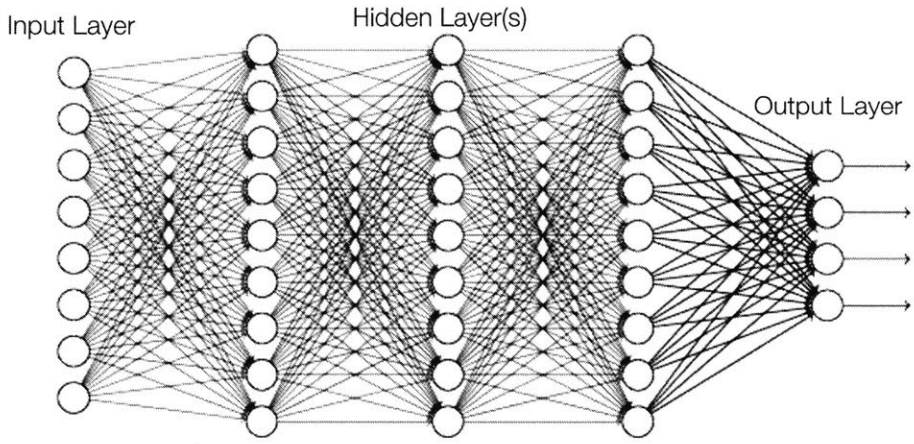

Figure 3-4: Artificial Neural Network

matching outputs. After "forward-propagating" an input thought an ANN, the difference between the ANNs estimated system output value and the true output value are compared. The residual (error) between these values is computed and "back-propagated" (not to be confused with optical back-propagation) through the ANN [30]. For each layer of the ANN, the error value is analyzed and used to adjust neuron weights. In this manner, successive examples (comprised on input-output pairs) "train" a neural network to obtain the correct output for its given inputs.

Deep Neural Networks (DNNs) are ANNs with many many layers that learn hierarchical representation of data. DNNs have been able to overcome previous issues with multilayer ANNs such as significantly increased training times and vanishing gradients by utilizing novel innovations in ANN connectivity such as: convolutions

[111, 109, 125, 98] for regularization and pruning in image recognition and classification tasks; nonlinearities, such as non-differentiable piecewise linear units [94] as opposed to the older sigmoidal functions that were differentiable but also prone to stagnation [96]; and algorithms, such as more efficient back propagation (backprop) [119, 112].

Recently, DNNs have seen wide usage in diverse domains: playing complex games on Atari 2600 [114] and Go [121]; object generation [92]; object detection [110]; and image processing: colorization [89], deblurring [90, 130, 124], and in-painting [129]. In Chapter 4, we propose and demonstrate that deep neural networks may "learn" to approximate solutions to inverse problems in computational imaging.

Most deep neural networks for image processing (regression, segmentation, or classification) are Convolutional Neural Networks (CNNs). In a CNN, inputs are passed from nodes of each layer to next, with adjacent layers connected by convolution, pooling, and non-linearity-generating operations. The convolution+downsampling steps of CNNs enable compressed representations of images (which in turn enable deeper networks and faster training) by combining spatially correlated features within images.

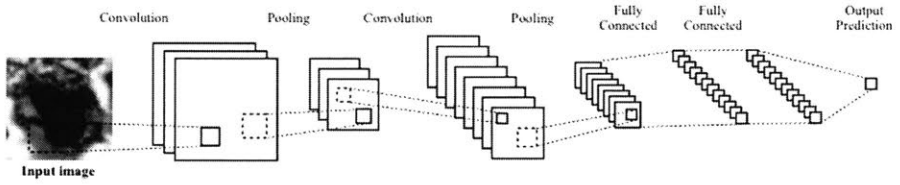

Figure 3-5: CNN for classification. At the very last layer of the CNN, the values represent an input's classification state likelihood.

A basic/common DNN architecture is "LeNet" [31] which utilizes several "convolution-downsample-ReLU" blocks sequentially followed by a fully connected layer before final output prediction. It was introduced by LeCun et al. in 1998 and was used primarily for optical character recognition [31].

Recently developed architectures include: ResNets [98] and DenseNets [99], and optimal/novel neural network architectures are an active field of study.

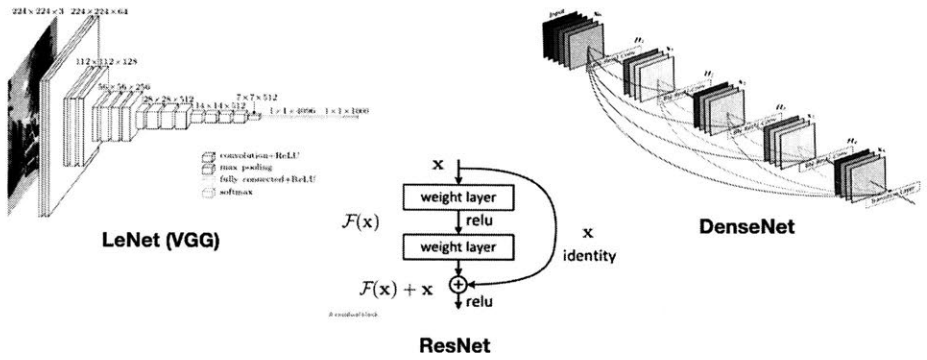

Figure 3-6: LeNet, ResNet, and DenseNet are common DNN architectures.

In a convolutional neural network (CNN), inputs are passed from nodes of each layer to the next, with adjacent layers connected by convolution, pooling, or non-linearity-generating operations. Convolutional ResNets extend CNNs by adding short term memory to each layer of the network. The intuition behind ResNets is that one only wants to add a new layer if you can get something extra out of adding that layer. ResNets ensure that the (N + 1)th layer learns something new about the network by also providing the original input (i.e., without any transformation performed) to the output of the (N + 1)th layer and performing calculations on the residual of the two. This forces the new layer to learn something different from what the input has already encoded/learned [98]. A single residual connection (representing the basic building block of a ResNet) is shown in the middle of figure 3-6.

DenseNets extend ResNets but connecting multiple layers within "dense blocks" to each other [99]. The image on the left of figure 3-6 illustrates a 5-layer dense block. Each layer within this dense block is directly connected to every subsequent layer in the block and is thus able to pass its feature maps directly to the remaining layers. The benefit of this architecture is that it significantly reduces the vanishing gradient problem faced by deep neural networks.

Chapter 4

Regularized Phase Space Imaging Using Sparsity Priors

We begin this chapter by reviewing phase-space representation of light fields and describing how phase-space tomography, ptychography, and Fourier ptychography can be represented in phase-space. Armed with this insight, we then propose and implement two algorithms utilizing low-rank matrix recovery (LRMR) in order to improve the robustness of two phase-space imaging techniques: phase-space tomography (PST) and Wigner distribution deconvolution (WDD).

4.1 Phase-space in optics

The "phase-space" (or "Wigner Distribution") of a light field describes the localized spatial frequencies (light ray directions) of light at any point in space. This spatial spectrogram can be more readily understood through an analogy with a musical score, which describes the localized *temporal* frequencies at any point in time (temporal spectrogram). Windowing temporal frequencies locally provides users the benefit of seeing local properties of the music (i.e., each note played) instead of only capturing global frequencies provided by Fourier Transforming the entire signal. Optical phase-space representations are "musical scores" for light distributions and provide valuable localized information about what's happening to light at any point in space.

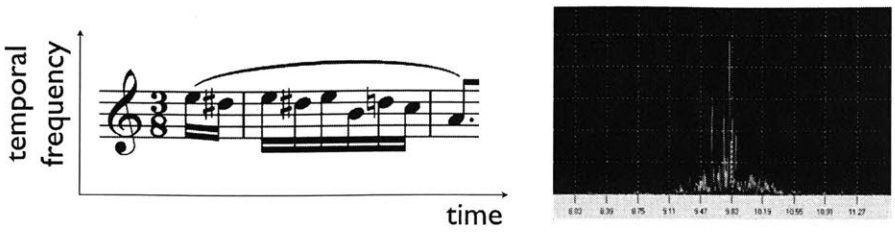

Figure 4-1: A temporal spectrogram (musical score) provides more insight into audio signal than a Fourier Transform of the entire signal can

Beyond providing localized insight into properties of light, phase-space representations of light are also extremely useful since [32]: (a) propagation of light in phase-space is a simple shear/rotation (with far-field propagation resulting in a 90 degree rotation), (b) measurement of light in phase-space is a simple projection, and (c) 4D phase-space is related to the 4D mutual intensity of an object via a 2D Fourier Transform. In the case of stationary quasi-monochromatic partially coherent light, this mutual intensity function describes all two-point correlation pairs and provides a complete characterization of the wave field.

Under the phase-space imaging framework, Phase-space tomography can be viewed as capturing lines in phase-space in a tomographic fashion and ptychography and Fourier ptychography can be seen as filling out phase-space with identical overlapping windows.

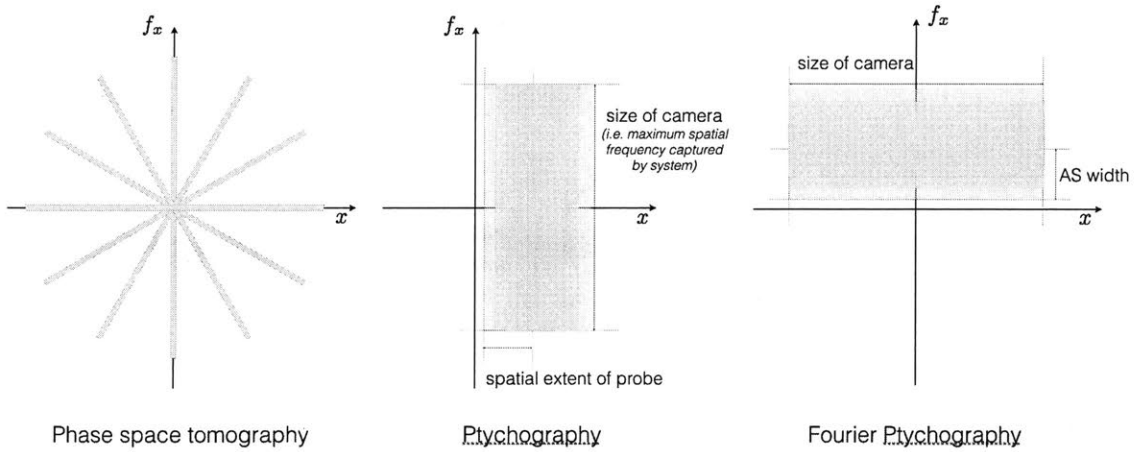

Figure 4-2: Phase-space tomography, ptychography, and Fourier ptychography in Phase-space

Phase-space tomography captures lines in phase-space via repeated rotations and projection (which fill out a 90 degree rotated phase-space/Ambiguity Function via the Fourier Slice Theorem). A single measurement in ptychography gives you all the frequency information about a localized area of the object (windowed by the probe). Shifting the probe moves the area of observation in the $\pm x$ direction in phase-space (with overlap). Likewise, a single measurement in Fourier ptychography gives you all the spatial information about a localized band in frequency space (windowed by the aperture stop of the system), and shifting the illumination angle moves the object frequencies passed through the imaging system AS in the $\pm f_x$ direction in phase-space.

The extended Ptychograhic Iterative Enginer (ePIE) [133] used to reconstruct ptychographic and Fourier ptychographic images enforces consistency in regions of phase-space overlap sequentially in order to iteratively solve for phase. However, as an iterative algorithm, it does not explicitly utilize the phase-space sampling geometry. In contrast, Wigner Distribution Deconvolution directly inverts the "convolution with window" operation in phase space.

4.2 Phase-space tomography

Phase space tomography estimates correlation functions entirely from snapshots in the evolution of the wave function along a time or space variable. In contrast, traditional interferometric methods require measurement of multiple two-point correlations. However, as in every tomographic formulation, undersampling poses a severe limitation. Here, we propose a compressive reconstruction of the classical optical correlation function (i.e. the mutual intensity function) using phase-space tomography and low-rank matrix recovery (LRMR). The LRMR algorithm makes explicit use of the physically justifiable assumption of a low-entropy source (or state).

The following subsections are adapted excepts from the published paper: "Experimental compressive phase space tomography" [33]. The author of this thesis contributed to original idea and initial simulations but not the experiments in this

paper.

4.2.1 Introduction

Correlation functions provide complete characterization of wave fields in several branches of physics, e.g. the mutual intensity of stationary quasi-monochromatic partially coherent light [34], and the density matrix of conservative quantum systems (i.e., those with a time-independent Hamiltonian) [35]. Classical mutual intensity expresses the joint statistics between two points on a wavefront, and it is traditionally measured using interferometry: two sheared versions of a field are overlapped in a Young, Mach-Zehnder, or rotational shear [36, 37] arrangement, and two-point ensemble statistics are estimated as time averages by a slow detector under the assumption of ergodicity [34, 38].

As an alternative to interferometry, phase space tomography (PST) is an elegant method to measure correlation functions. In classical optics, PST involves measuring the intensity under spatial propagation [39, 40, 41] or time evolution [42]. In quantum mechanics, analogous techniques apply [43, 44, 45, 46]. However, the large dimensionality of the unknown state makes tomography difficult. In order to recover the correlation matrix corresponding to just n points in space, a standard implementation would require at least n^2 data points.

Compressive sensing [47, 48, 49] exploits sparsity priors to recover missing data with high confidence from a few measurements derived from a linear operator. Here, sparsity means that the unknown vector contains only a small number of nonzero entries in some specified basis. Low-rank matrix recovery (LRMR) [50, 51] is a generalization of compressive sensing from vectors to matrices: one attempts to reconstruct a high-fidelity and low-rank description of the unknown matrix from noisy or incomplete measurements.

It is worth noting that LRMR came about in the context of compressive quantum state tomography (QST) [52], which utilizes different physics to attain the same end goal of reconstructing the quantum state. In PST, one performs tomographic projection measurements, rotating the Wigner space between successive projections

by evolving the wave function [39, 40]. This is directly analogous to the classical optical experiment we are presenting here, where we perform intensity measurements (i.e., tomographic projections in Wigner space) and utilize propagation along the optical axis to rotate the Wigner space between projections. The difference lies in the fact that in QST the state is recovered via successive applications of the Pauli dimensionality-reducing operator, and there is no need to evolve the state. Nevertheless, both approaches lead to the same Hermitian LRMR problem, as long as the assumption of a quasi-pure unknown state is satisfied.

In [53], it was shown that estimation of a low-rank matrix of dimension n and rank r requires only $O(rn\ln n)$ to $O(rn\ln^2 n)$ data points. A similar LRMR method was also used to recover the complex amplitude of an unknown object under known illumination [54, 55, 56]. Since the complex amplitude of the object is time-invariant, a rank-one solution was assumed in these works.

The low-rank assumption for classical partially coherent light anticipates a source composed of a small number of mutually incoherent effective sources, i.e. coherent modes [57], to describe measurements. This is essentially equivalent to the low entropy assumption [52], i.e. a nearly pure quantum state in the quantum analogue. This assumption is valid for lasers, synchrotron and table-top X-ray sources [58], and Kohler illumination in optical microscopes [34]. An additional requirement for LRMR to succeed is that measurements are "incoherent" with respect to the eigenvectors of the matrix, i.e. the measured energy is approximately evenly spread between modes [53, 54]. Diffraction certainly mixes the coherent modes of the source rapidly, so we expect LRMR to perform well for classical PST. The same expectation for QST has already been established [52].

4.2.2 Theory and Method

The two-point correlation function of a stationary quasi-monochromatic partially spatially coherent field is the mutual intensity function [34]

$$J(x_1, x_2) = \langle g^*(x1)g(x2)\rangle,$$

where $\langle\cdot\rangle$ denotes the expectation value over a statistical ensemble of realizations of the field $g(x)$. The measurable quantity of the classical field, i.e. the intensity, after propagation by distance z is [34]:

$$I(x_0;z) = \iint dx_1 dx_2 J(x_1,x_2)\exp\left(-\frac{i\pi}{\lambda z}(x_1^2 - x_2^2)\right)\exp\left(i2\pi\frac{x_1 - x_2}{\lambda z}x_0\right).$$

This can be expressed in operator form as:

$$I = \mathrm{Tr}(P_{x_0} J)$$

where P denotes the free-space propagation operator that combines both the quadratic phase and Fourier transform operations in Eq. (2), $\mathrm{Tr}(\cdot)$ computes the trace, and x_0 denotes the lateral coordinate at the observation plane. By changing variables $x = (x_1 + x_2)/2$, $x' = x_1 - x_2$ and Fourier transforming the mutual intensity with respect to x, we obtain the Ambiguity Function (AF) [59, 60, 61]:

$$\mathcal{A}(u',x') = \int J\left(x + \frac{x'}{2}, x - \frac{x'}{2}\right)\exp(-i2\pi u' x)dx$$

Eq 2. can be written as [6-8,26,28]:

$$\tilde{I}(u';z) = (A)(u', \lambda z u')$$

where $\tilde{I}$ is the Fourier transform of the vector of measured intensities with respect to x_0. Thus, radial slices of the AF may be obtained from Fourier transforming the vectors of intensities measured at corresponding propagation distances, and from the AF the mutual intensity can be recovered by an additional inverse Fourier transform, subject to sufficient sampling. To formulate a linear model for compressive PST, the measured intensity data is first arranged in Ambiguity space. The mutual intensity is defined as the "sparse" unknown to solve for. To relate the unknowns (mutual intensity) to the measurements (AF), the center-difference coordinate-transform is first applied, expressed as a linear transformation T upon the mutual intensity J,

followed by Fourier transform F, and adding measurement noise e as:

$$\mathcal{A} = \mathcal{F} \cdot T \cdot J + e$$

The mutual intensity propagation operator is unitary and Hermitian, since it preserves energy. We use eigenvalue decomposition to determine the basis where the measurement is sparse. The resulting basis, i.e. the set of eigenvectors, is also known as coherent modes in optical coherence theory, whereas the whole process is known as coherent mode decomposition [57]. The goal of the LRMR method is to minimize the number of coherent modes to describe measurements. By doing LRMR, we impose two physically meaningful priors: (1) existence of the coherent modes [57], and (2) sparse representation of the partially coherent field in terms of coherent modes

Mathematically, if we define all the eigenvalues λ_i and the estimated mutual intensity as $\hat{J}$, the method can be written as:

$$\begin{aligned} \text{minimize} \quad & \text{rank}(\hat{J}) \\ \text{subject to} \quad & \mathcal{A} = \mathcal{F} \cdot T \cdot \hat{J} \\ & \lambda_i \geq 0, \text{ and } \sum_i \lambda_i = 1 \end{aligned} \tag{4.1}$$

Direct rank minimization is NP-hard; however, it can be accomplished by solving instead a proxy problem: convex minimization of the "nuclear norm" (L_1 norm) of the matrix J [50, 27]. The corresponding problem is stated as:

$$\begin{aligned} \text{minimize} \quad & \|J\|_* \\ \text{subject to} \quad & \mathcal{A} = \mathcal{F} \cdot T \cdot \hat{J} \\ & \lambda_i \geq 0, \text{ and } \sum_i \lambda_i = 1 \end{aligned} \tag{4.2}$$

where the nuclear norm is the sum of the singular values $\sigma_i = |\lambda_i|$, $\|J\|_* = \sum_i \sigma_i$. This problem is convex and a number of numerical solvers can be applied to solve it. In our implementation, we used the singular value thresholding (SVT) method [71]. The output estimate after each iteration of SVT typically has a sub-normalized total

energy, i.e. $\sum_i \lambda_i < 1$; we compensated for this by renormalizing at the end of each iteration [52].

4.3 Denoised Wigner distribution deconvolution via low-rank matrix completion

Wigner Distribution Deconvolution (WDD) is a decades-old method for recovering phase from intensity measurements. Although the technique offers an elegant linear solution to the quadratic phase retrieval problem, it has seen limited adoption due to its high computational/memory requirements and the fact that the technique often exhibits high noise sensitivity. Here, we propose a method for noise suppression in WDD via low-rank noisy matrix completion. Our technique exploits the redundancy of an object's phase space to denoise its WDD reconstruction. We show in model calculations that our technique outperforms other WDD algorithms as well as modern iterative methods for phase retrieval such as ptychography. Our results suggest that a class of phase retrieval techniques relying on regularized direct inversion of ptychographic datasets (instead of iterative reconstruction techniques) can provide accurate quantitative phase information in the presence of high levels of noise.

The following subsections were adapted from the published paper: "Denoised Wigner distribution deconvolution via low-rank matrix completion" [85].

4.3.1 Introduction

Wigner Distribution Deconvolution was first proposed as a method for phase retrieval in 1989 by Bates and Rodenburg [62]. The technique is particularly elegant in that it solves the quadratic phase retrieval problem using only linear computations. Moreover, unlike other techniques for phase retrieval which rely on systems with coherent illumination (or at least which are assumed to be approximately coherent), WDD can be adapted for use in experimental setups utilizing partially coherent illumination [63].

Wigner Distribution Deconvolution has been demonstrated successfully previously

in the optical regime [64], in the X-ray regime [65], and with electrons [66]. Although these original experiments occurred over two decades ago, WDD has failed to gain momentum as a technique in large part due to the following reasons: (1) the sheer volume of data required for WDD is enormous (a 4-D set of data must be collected to reconstruct a 2-D image), (2) the deconvolution step in WDD is often ill-conditioned, resulting in local amplification of error in the deconvolution process, (3) the rise of simpler iterative techniques for phase retrieval like ptychography.

Computational power has increased dramatically since the 1990's, reducing some of the first restraint. Recently, the authors in [67] proposed a "projection strategy", an optimized probe, and an iterative "Wigner replacement" procedure for noise suppression in WDD. In this paper, we propose and demonstrate a denoising strategy that exploits the low-rank nature of recovered mutual intensity in WDD. We show that our new technique outperforms traditional WDD, the WDD "projection+replacement strategy" proposed in [67], as well as conventional iterative methods like ePIE (the extended Ptychographical Iterative Engine) in the presence of significant levels of noise.

4.3.2 Wigner distribution deconvolution

An example experimental arrangement for WDD is shown in Fig. 4-3. In order to provide a more intuitive understanding of WDD, the illustrations and equations in Sections 1-4 are done with a 1-dimensional object and probe (each of which have a 2-D Wigner space) under coherent illumination. The equations derived and used in these sections are extendible to 2-dimensions (4-D Wigner space), and in section 5, simulations are done in 2-D/4-D.

In WDD, an object $o(x)$ is illuminated by a shifted probe $p(x - s_i)$, and the resultant field $o(x)p(x - s_i)$ is measured in the Fourier plane as an intensity (time-average). Collecting all of the intensities measured over all of the shifts s yields the

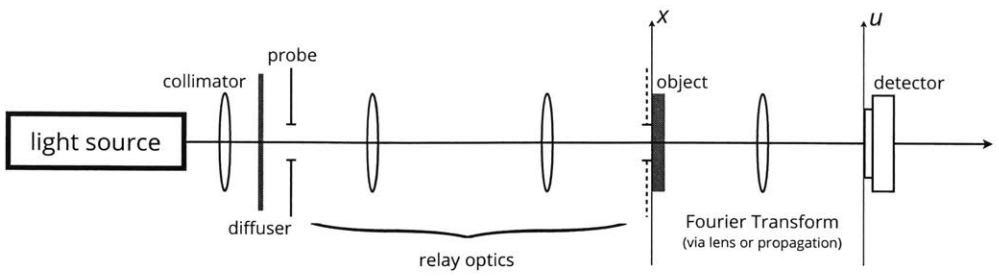

Figure 4-3: Experimental setup for WDD. A shifted probe is imaged onto the object plane and the resultant field $p(x - s_i) \cdot o(x)$ is propagated to a Fourier plane where its intensity is measured.

2-D matrix (4-D if the objects 2-D):

$$I_{\text{out}}(u, s) = \left| \int p(x - s) \cdot o(x) e^{-j2\pi xu} dx \right|^2 \tag{4.3}$$

where x represents the coordinates at the object plane, u represents the coordinates at the detector plane (spatial frequency), and s represents the lateral displacement of the probe. Applying a 1-D Fourier Transform in s and a 1-D Inverse Fourier Transform in u to $I(u, s)$ yields [63]:

$$H(x, s') = \iint p^*(t) \cdot p(t + x) \cdot o^*(x_1) \cdot o(x_1 + x)\, e^{-j2\pi(x_1 - t)s'} dx_1 dt \tag{4.4}$$

If s is taken as a cyclical coordinate, then the output $H(x, s')$ is sampled in s' (Fourier Series) and the integral above becomes separable [67], yielding:

$$H(x, s') = \underbrace{\int p^*(t) \cdot p(t + x) e^{j2\pi ts'} dt}_{\mathcal{W}_{\text{probe}}(x, -s')} \cdot \underbrace{\int o^*(x_1) \cdot o(x_1 + x)\, e^{-j2\pi x_1 s'} dx_1}_{\mathcal{W}_{\text{object}}(x, s')} \tag{4.5}$$

Although the step from Eq. (4.4) to Eq. (4.5) assumes s is a cyclical coordinate, the probe displacement is actually linear. However, as noted in [67], the dataset generated from linear and cyclical scans are equivalent provided that $\Delta_x \geq \Delta_{x,\text{obj}} + \Delta_{x,\text{pr}}$, where Δ_x is the lateral extent of the reconstruction, and $\Delta_{x,\text{obj}}$, $\Delta_{x,\text{pr}}$ are the

sizes of the support region of the object and probe, respectively. If the object or probe are not strictly finite (e.g., if they tail off), then a window function may be employed as shown in [64] to meet the above criterion.

The authors note here that although the original WDD derivation was derived with continuous variables [63], in practice Fourier computations are done on a computer where signals are discretized and a periodic extension is assumed/applied. For these signals (discretized by pixel size δx, and periodically extended with periodicity equal to the length of the recorded signal Δx), a Discrete Fourier Transform (DFT) or Inverse Discrete Fourier Transform (IDFT) is applied to in lieu of the Fourier Transform (FT) or Inverse Fourier Transform (IFT) into order to convert from a sampled periodic signal in the spatial domain (with spacing δx and periodicity Δx) to or back from a sampled periodic signal in the frequency domain (with spacing $\frac{1}{\Delta x}$ and periodicity $\frac{1}{\delta x}$). In this paper we use the terminology FT/IFT when discussing theory and DFT/IDFT when discussing implementation.

In Eq. (4.5) above, the two terms in underbraces are the Wigner distributions of the probe and object. Wigner distributions can be thought of as self-windowed spectrograms representing the object/probe's phase-space (local spatial frequency spectra) [82]. Applying a 1D Fourier Transform to an object's Wigner distribution yields the sheared mutual intensity of that object (or of that object's Fourier Transform). The mutual intensity of any two fields represents all of the two-point correlations between every possible point-pair between the fields, and in the discretized coherent case is simply the outer product (rank $r = 1$) between the object and its conjugate, $o^*(x_1)o(x_2)$.

From Eq. (4.5), one can obtain $\mathcal{W}_{\text{object}}(x, s')$ by dividing $H(x, s')$ by the Wigner distribution of the probe. To avoid errors introduced due to division by zero, a Tikhonov (Wiener) filter is commonly used to regularize the inversion. Thus, the object's estimated Wigner distribution is calculated as:

$$\mathcal{W}_{\text{object}}(x, s') = \frac{\mathcal{W}^*_{\text{probe}}(x, -s')H(x, s')}{|\mathcal{W}_{\text{probe}}(x, -s')|^2 + \epsilon} \tag{4.6}$$

An optimal spectral filter would regularize the inversion in Eq. (4.6) based on knowledge of the noise spectrum. However, since the recorded intensities are corrupted by both Poisson and Gaussian noise, the optimal Wiener filter is object-dependent and not easily determinable. Maximum-likelihood estimators have been employed by others to deal with mixed Poisson and Gaussian noise in ptychographic reconstruction [80, 81, 84], but in this paper we denoise our WDD reconstruction via Noisy Low-Rank Matrix Completion performed at a later step. Thus, the filter above assumes a flat noise spectrum, does not attempt to denoise, and is employed only to regularize the inversion/deconvolution (i.e., to deal with its ill-conditionedness at low $\mathcal{W}_{\text{probe}}$ values).

To retrieve the object $o(x)$ from its Wigner distribution $W_{\text{obj}}(x, s')$, we may either:

1. apply a 1-D IDFT in s' to $\mathcal{W}_{\text{obj}}(x, s')$, yielding: $o^*(x_1)o(x_1 + x)$, (the sheared mutual intensity of the object), or

2. apply a 1-D DFT in x to $\mathcal{W}_{\text{obj}}(x, s')$, yielding: $O(u)O^*(u - s')$ (the sheared mutual intensity of the object's frequency spectrum).

In the absence of noise and given complete coverage of the object's mutual intensity, $o(x)$ can be obtained directly from the sheared mutual intensity by unshearing it and performing a singular value decomposition (SVD) or an eigendecompostion. Exactly one singular value should exist, and the corresponding left singular vector should yield $o(x)$ (scaled by $\sqrt{\sigma_1}$). Likewise, an SVD of the unsheared $O(u)O^*(u+s')$ should yield $O(x)$, which can be inverse Fourier transformed to give $o(x)$.

It should be noted here that Wigner Distribution Deconvolution and applications of PhaseLift [78, 83], both of which analyze ptychographic data in a higher dimension for phase retrieval, differ in terms of fundamental approach: WDD arranges the data in a way such that one can retrieve complex amplitude-phase via linear computations while PhaseLift changes the quadratic equations (one for each measurement) into a Second Order Cone Problem (SOCP) via a lifting procedure. As opposed to getting the higher dimensionality via lifting of each underlying measurement equation, WDD gets its higher dimensionality by putting all the measurements into a square matrix,

$I(u, s)$ with position along one coordinate and shift along the other. In the algorithm presented in this paper, Low-Rank Matrix Recovery/Completion (LRMR/LRMC) is used to de-noise and isn't a fundamental requirement of WDD theory (e.g., in the case where the probe isn't band-limited, our approach would simply reduce to doing an SVD); *the WDD formulation is entirely linear.* In contrast, using PhaseLift, even if one had a perfect probe, she/he would still have to solve a SOCP.

In practice, probes are not perfect, and complete coverage of the object's mutual intensity is not attainable due to the finite extent or finite bandwidth of the probe. Fig. 4-4 illustrates the problem that arises in these situations by examining the Wigner Distributions of the probe. In the central column of Fig. 4-4, the probe is limited in lateral extent, limiting the reconstructable $\mathcal{W}_{\text{obj}}$ in the x direction. In the column on the right, the probe is band-limited, limiting the reconstructable $\mathcal{W}_{\text{obj}}$ in the s' direction.

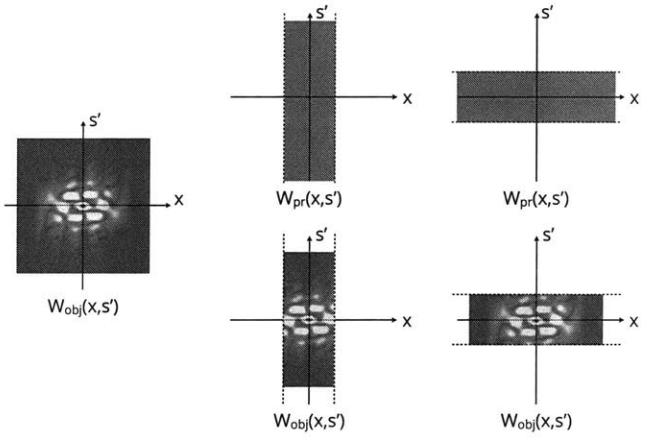

Figure 4-4: Effects of a probe with limited spatial extent and limited bandwidth on object reconstruction in WDD. (Left) Actual Wigner distribution of object. (Top Center) Wigner distribution of a finite-extent probe. (Top Right) Wigner distribution of a band-limited probe. (Bottom Center) Recovered Wigner distribution of object using finite-extent probe. (Bottom Right) Recovered Wigner distribution of object using band-limited probe.

From the diagrams in Fig. 4-4, we observe that if the spatial extent of the probe is less than that of the object (central column), then:

- Technique 1 (1-D IDFT in s') gives an accurate mutual intensity over the points that it covers but fails to cover all points.

- Technique 2 (1-D DFT in x) fails to return an accurate mutual intensity.
  - Fourier Transforming W_{obj} in x_1 results in a blurring of $O(u)O^*(u - s')$ along u. It may be possible to retrieve an accurate estimate of $O(u)$ from the unidirectionally-blurred mutual intensity matrix, and this is a topic of interest for future work.

If the bandwidth of the probe is less than that of the object (right column), then:

- Technique 1 fails to return an accurate mutual intensity (result is unidirectionally-blurred).

- Technique 2 gives an accurate mutual intensity over the points that it covers but fails to cover all points.

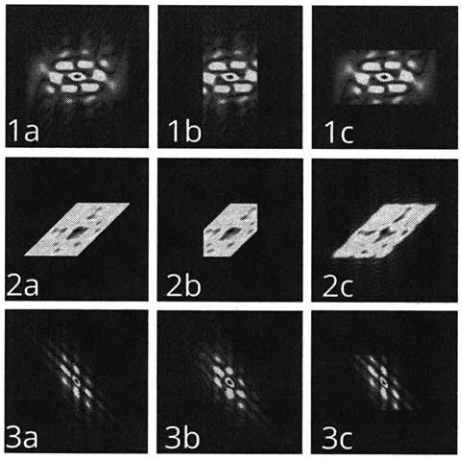

Figure 4-5: Effects of a probe with limited spatial extent and limited bandwidth on mutual intensity reconstruction in WDD. (1) Wigner distribution of object: (a)actual (b)retrieved using probe with limited spatial extent, (c)retrieved using probe with limited bandwidth. (2) sheared mutual intensity of object, $o^*(x_1)o(x_1 + x)$: (a)actual (b)retrieved if probe has limited spatial extent, (c)retrieved if probe has limited bandwidth, (3) sheared mutual intensity of object's Fourier Transform, $O(u)O^*(u - s')$: (a)actual (b)retrieved if probe has limited spatial extent, (c)retrieved if probe has limited bandwidth,

The mutual intensities obtained in the four cases above are illustrated in Fig. 4-5. In the cases where an accurate but incomplete mutual intensity is retrieved (cases 2b and 3c in Fig. 4-5), several methods have been proposed to recover the

amplitude and phase of the corresponding object. Bates and Rodenburg used a "stepping out" method in their original work on WDD [63], and more recently Li, Edo, and Rodenburg utilized a "projection strategy" [67]. In the absence of noise, both the "stepping out" and "projection" methods can accurately reconstruct an object from its partially measured mutual intensity. However, the "stepping out" method cannot make use of all available data simultaneously and suffers from accumulation of error [67]. The "projection method" has been shown to be more robust to noise due to its utilization of more data points, but each projection "los[es] the ... highest frequencies" [67]. Noise reduction is attained by averaging across multiple datapoints, and the projections themselves does not attempt to separate noise from underlying signal.

4.3.3 Noise considerations in WDD

Common sources of noise in optical imaging systems include: shot noise, dark current noise, and Johnson noise. Shot noise (Poisson noise) is inherent in any imaging system due to the collection of a finite number of photons per unit time. Shot noise is correlated with the intensity of each pixel, and we apply it to our simulated data. Johnson noise (Gaussian noise) occurs due to thermally-induced current fluctuations in the load resistor. Johnson noise is independent of the original intensity at each pixel, and we add it to our simulated data. Dark current bias and noise arised from thermally-induced free electron production on the detector. Since dark current bias can be addressed by dark-frame subtraction, we do not apply it in our simulated data. Dark current noise is Poisson in nature and we apply it to our simulated data.

Combining Poisson and Gaussian noise sources into one noise term, noise propagation in WDD can be seen as:

$$I(u,s)+\mathcal{N}(u,s) \xrightarrow{\text{2-D DFT}} H(x,s')+\widetilde{\mathcal{N}}(x,s') \xrightarrow{./\mathcal{W}_{\mathrm{p}}(x,s')} \mathcal{W}_{\mathrm{o}}(x,s')+\frac{\widetilde{\mathcal{N}}(x,s')}{\mathcal{W}_{\mathrm{p}}(x,s')} \quad (4.7)$$

At locations where $\mathcal{W}_{\mathrm{probe}}(u,s)$ has very low values, the inversion of Eq. (4.6) is ill-conditioned. Any noise at these locations in $H(u,s)$ are amplified, resulting in an

inaccurate estimate of the object's Wigner distribution, $\mathcal{W}_{\text{object}}(u, s)$.

An ideal probe would have a Wigner distribution with a perfectly flat magnitude. Although one cannot create such a probe, a diffuser offers a close approximation (yielding few local minima) [68, 69]. For simulations in this paper, a model diffuser probe is used.

While diffuser probes can be used in experimental setups, it is difficult to accurately measure such a probe beforehand. Projection-based ptychography solvers (e.g., the Ptychographical Iterative Engine/PIE algorithm [62]) address this problem by including a probe recovery step in their iterative loop, solving for both the unknown object and probe simultaneously. While we do not explore such a possibility here, it is possible that alternating between probe and object recovery in our WDD algorithm with an outer loop (in a manner akin to the ePIE algorithm) could allow for simultaneous object and probe recovery. Also, it may be possible to employ a hybrid approach, using an ePIE-like algorithm to first get an approximate probe estimate followed by application of our algorithm for object recovery.

As discussed in Section 2, earlier methods of object retrieval from retrieved partial mutual intensity in WDD are inherently noise-limited due to the fact that they do not utilize all of the available data and that they do not make any effort to separate the noise from the signal. The authors in [67] utilized "iterative replacement" (in conjunction with their "projection" technique) to help denoise the object's estimated Wigner Distribution, but this method of denoising is suboptimal and in our simulations use of iterative replacement did not significantly improve results. As mentioned earlier, if the entirety of the mutual intensity matrix is available (albeit potentially noisy), then one may retrieve $o(x)$ via singular value decomposition. This method has the benefit of being optimal in the presence of Gaussian noise.

Motivated by this observation, we propose matrix completion as a viable alternative to the "stepping out" and "projection" methods for retrieving $o(x)$ or $O(u)$ from a noisy and incomplete $o^*(x_1)o(x_1 + x)$ or $O(u)O^*(u - s')$, respectively. In doing so, we cast the issue of completing and denoising the mutual intensity matrix as a matrix completion problem in the presence of noise. The underlying matrix (mutual inten-

sity) is known to be low-rank (in the perfectly coherent case, rank $r = 1$), suggesting that such a technique may work well, even in the presence of high levels of noise and incomplete measurement of entries.

4.3.4 Noisy matrix completion

In the "*noisy matrix completion problem*," one tries to recover an underlying low-rank matrix M from noisy and incomplete measurements of its entries. M can by any real or complex matrix with dimensions $m \times n$ and rank $r \ll m, n$. In our WDD formulation, M is the unsheared mutual intensity matrix, $o(x_1)^* o(x_2)$ or $O(u_1)^* O(u_2)$, but in general the matrix M need not be symmetric. Such a matrix has a (non-unique) singular value decomposition:

$$M = U \Sigma V^T$$

Additive or applied noise can be modeled as perturbations to each matrix entry such that:

$$\tilde{M}_{ij} = M_{ij} + Z_{ij} \tag{4.8}$$

where the matrix Z accounts for the added/applied noise and the matrix $\tilde{M}$ is the noisy "approximately low-rank" matrix that is (partially) observed/measured. Of the complete matrix $\tilde{M}$, only a subset Ω of its entries are actually observed. The objective of *noisy matrix completion* is to find a low-rank estimation X of the original matrix M based on the observed indices Ω of the noisy matrix $\tilde{M}$.

In the case where there is no noise ($\tilde{M} = M$), the following optimization may be used to recover the low-rank matrix exactly, provided that $|\Omega|$ is large enough and satisfies certain "incoherence" conditions [77]:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \text{rank}(X) \\ \text{subject to} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \end{aligned} \tag{4.9}$$

where $\mathcal{P}_\Omega(\cdot)$ is the operator selecting which entries of the matrix we have measured

and is defined as:

$$\mathcal{P}_\Omega(M)_{ij} = \begin{cases} M_{ij}, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

Since the problem in (4.9) is NP-hard, a common approach is to apply a convex relaxation [77, 27, 72] which turns the problem into:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \|X\|_* \\ \text{subject to} \quad & \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M) \end{aligned} \tag{4.11}$$

where $\|X\|_*$ denotes the nuclear norm of X (i.e., the sum of all the singular values of X).

Adding noise (which may be adversarial/worst-case), the problem in (4.11) becomes [27, 74, 75]:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \|X\|_* \\ \text{subject to} \quad & \|\mathcal{P}_\Omega(X - \tilde{M})\|_F \leq \delta \end{aligned} \tag{4.12}$$

which can be alternatively expressed in Lagrangian form as:

$$\text{minimize} \quad \mu\|X\|_* + \frac{1}{2}\|\mathcal{P}_\Omega(X - \tilde{M})\|_F^2 \tag{4.13}$$

If an upperbound on the rank of X is known, then one may solve the related rank-r matrix approximation problem [73, 76]:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \|\mathcal{P}_\Omega(X - \tilde{M})\|_F \\ \text{subject to} \quad & \text{rank}(X) \leq r \end{aligned} \tag{4.14}$$

If certain properties are known about the noise, then we may change the constraints in the above equations to reflect our apriori information. For example, if the noise is impulsive (e.g., dead pixels on a camera), then one may use the entrywise L_1 norm in place of the Frobenius norm in the equations above ($\|\mathcal{P}_\Omega(X - Y)\|_F \rightarrow \|\mathcal{P}_\Omega(X - Y)\|_{L1}$). Likewise, if the noise is known to be low-rank, then we may use

the nuclear norm in place of the Frobenius norm.

Notably, if all entries are known, then (4.14) reduces to classical PCA, which solves:

$$\begin{aligned} \underset{X_{ij}}{\text{minimize}} \quad & \|X - \tilde{M}\|_F \\ \text{subject to} \quad & \text{rank}(X) \leq r \end{aligned} \tag{4.15}$$

The authors in [27] and [72] have proven performance guarantees for matrix completion with noisy observations. Provided that the singular vectors of M are sufficiently spread out over the matrix ("strong incoherence property" [27]) and that the cardinality of the sampled set is at least the number of degrees of freedom times a few logarithmic factors, then the nuclear norm minimization is exact in the noiseless case and bounded in the noisy case – i.e., the root mean square error between the estimated (completed and denoised) X matrix and our actual matrix M is bounded and scales with the Frobenius norm of the error.

In this paper, we adapt a modified version of the Singular Value Thresholding (SVT) algorithm [27] for noisy matrix completion (to solve Eq. (4.14) above). Our adaptation alternates between: (1) replacing the "measured indices" ($\mathcal{P}_\Omega$) of the current best-estimate of the object's mutual intensity with the values of the measured object mutual intensity at those indices, (2) eigendecomposing the updated estimate for object estimated mutual intensity and keeping just the first eigenvector/value to obtain an estimate of the object, and (3) generating a new estimated object mutual intensity from the new object estimate. Solving the noisy matrix completion problem in this manner simultaneously completes and denoises the noisy (and incomplete) mutual intensity matrix X by enforcing a rank-1 constraint on the output of each iteration in the LRMC/LRMR loop. A flowchart of the algorithm is shown in Fig. 4-6.

Although the formulations of LRMC above do not place a positive semidefinite constraint on the matrix X, our algorithm returns an X that is by construction positive-semidefinite since it is generated from the outer product of a single eigenvector.

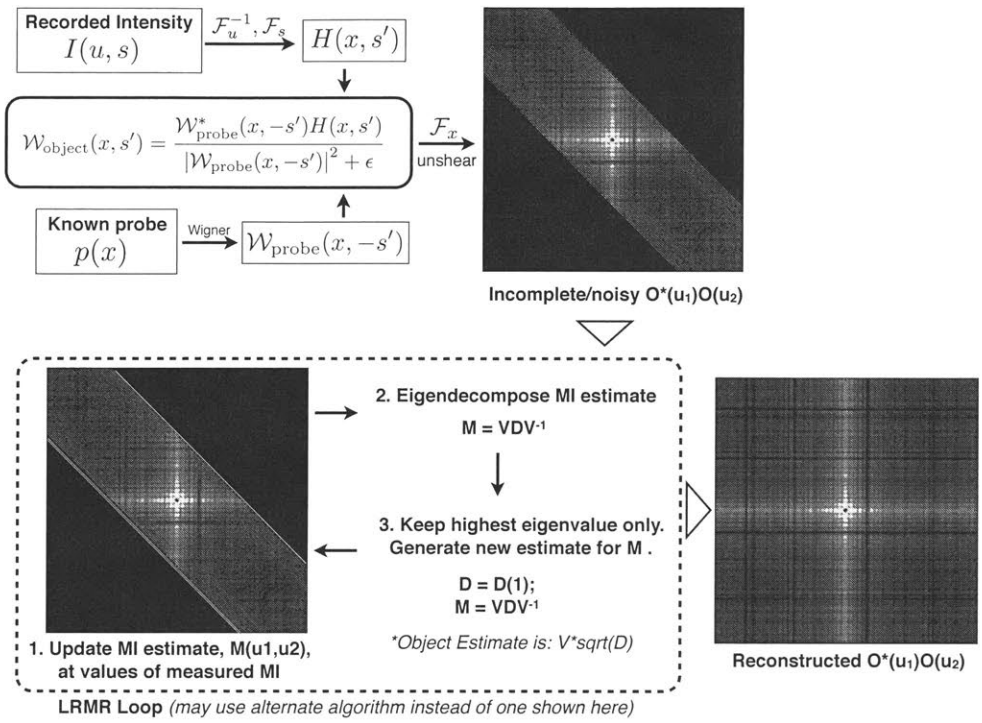

Figure 4-6: Flow chart of denoised WDD with LRMC using modified SVT. Application of our algorithm to a simulated noisy 1-D dataset with a band-limited probe yields a denoised rank-1 mutual intensity reconstruction similar to rightmost image (matrix is simultaneously completed and denoised).

4.3.5 Comparison of techniques for WDD

In this section, we apply our algorithm, the "projection method" algorithm from [67], the PIE (alternating projections) to a test object with variable amplitude and phase at varying noise levels. In our simulations, the calculation window is 63 pixels wide, with each pixel being square and $2.02\mu m$ wide in real space and $787m^{-1}$ wide in reciprocal space. The probe and object were roughly equivalent in size and each fill up around half of the calculation window (31 pixels). The probe was modeled as a diffuser imaged onto the object plane through a system with a low NA in order to better illustrate the "matrix completion" aspect of our algorithm (in addition to its denoising capability). Since the probe is band-limited and not strictly finite, we employ a windowing function as in [64].

The object used in our simulation is shown below in Fig. 4-7. For ease of visualization of phase error, we set a colormap range of $0 - \pi$ for display/comparison of phase

values, but we limited the range of our object's actual phase to between $\pi/4$ and $3\pi/4$ (such that estimated phase values slightly above or below the maximum/minimum phase values of the object won't cause wrapping visualization issues). The object's modulus ranges from $0 - 1$.

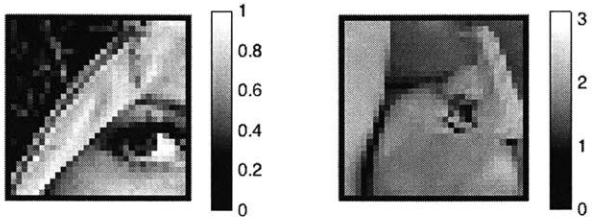

Figure 4-7: Original object's modulus (left) and phase (right)

A simulated diffraction dataset was generated using our probe and object, and we applied Poisson-distributed noise with total counts of $10^9, 10^8, 10^7, 10^6$, and 10^5 to that dataset. Reconstruction of $o(x)$ from these noisy datasets was performed using WDD with LRMC, WDD with the "projection method" + "replacement iterations", and ptychographic reconstruction via alternating projections.

The results of our simulations are shown in Fig. 4-8. Computations were performed on a 3.5 GHz 6-Core Intel Xeon processor (2013 Mac pro). Ptychographic reconstruction using the PIE took 7 seconds, reconstruction via the "projection method" from [67] took 55 seconds, and reconstruction via our LRMC algorithm took 162 seconds. For larger-sized problems, our algorithm may be parallelized to increase computation speed.

We performed our simulation 100 times with different simulated datasets (different diffuser + applied noise each iteration) and the average Mean-Squared-Error of each technique (and standard deviation over the trials) is shown in Table 4.1. At low photon counts (high noise), LRMC significantly outperforms both the "projection method" and PIE in our simulations. At high photon counts all three methods are competitive. However, as discussed in section 2, the projection method cannot recover the complete frequency spectrum of the object when the probe is band-limited.

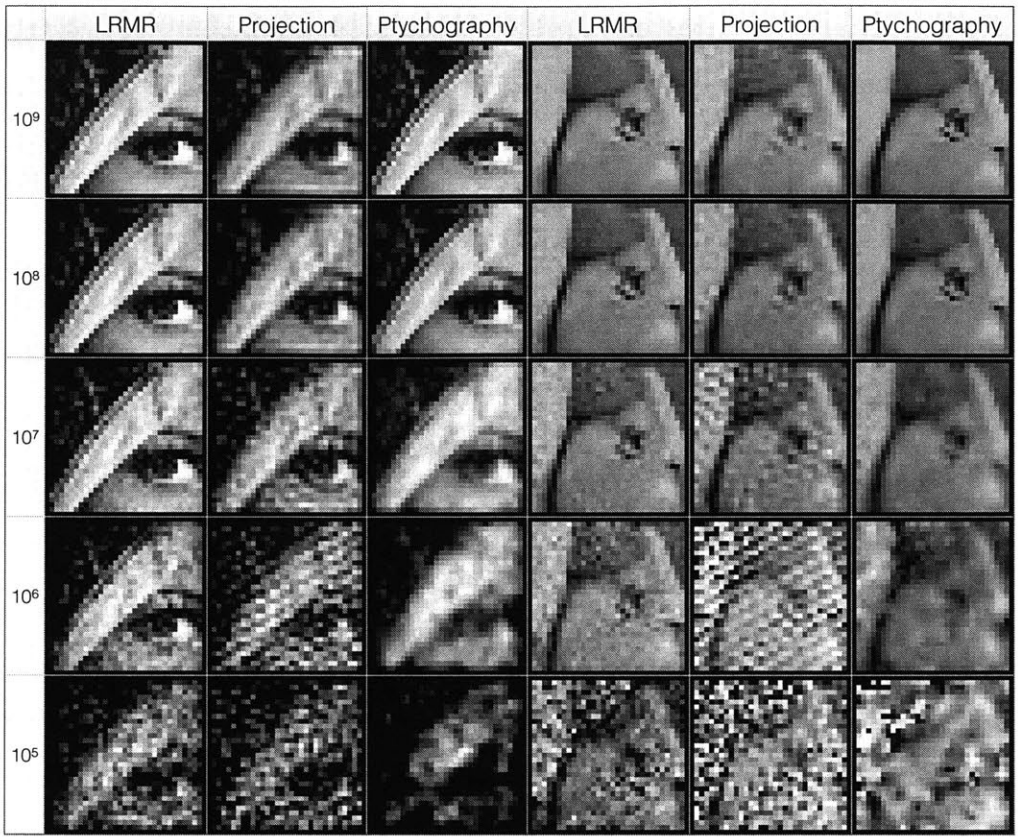

Figure 4-8: Results of WDD using: LRMR/LRMC, projection method from [67], and PIE at different noise levels. From left-to-right, the columns represent: moduli (1-3) of reconstructed object using LRMR/LRMC, "projection method", ptychography, and phase (4-6) of reconstructed object using LRMR/LRMC, "projection method", ptychography. The rows represent the total photon counts available to the detector in a shot-noise limited imaging system.

4.3.6 Discussion/Conclusion

In this paper, we have presented a method for noise suppression in WDD which exploits the redundancy of an object's phase space/mutual intensity to denoise (and complete if necessary) the WDD reconstruction. We show in model calculations that our technique outperforms other WDD algorithms as well as conventional iterative methods for phase retrieval such as ptychography.

Several iterative ptychographic reconstruction algorithms have been developed in recent years to take noise statistics into account [81, 84]. We have not compared our algorithm against these Maximum-Likelihood-based methods but note here that incorporation of noise-model priors is not mutually exclusive to our algorithm's uti-

Table 4.1: Mean Squared Error of object reconstructions from datasets with different noise levels over 100 trials.

Photon Count	LRMC	"Projection Method"	PIE
10^9	0.55 ± 0.05	1.26 ± 0.09	0.23 ± 0.04
10^8	1.3 ± 0.2	1.8 ± 0.1	0.8 ± 0.1
10^7	1.5 ± 0.2	2.6 ± 0.7	2.2 ± 0.2
10^6	3.7 ± 0.5	4.4 ± 1.3	5.4 ± 0.2
10^5	5.0 ± 1.2	6.6 ± 2.2	11.4 ± 0.5

lization of the known low-rank nature of the output mutual intensity for noise suppression. Thus, a hybrid algorithm could be developed which utilizes both low-rank and noise-model priors to retrieve the complex object in the presence of high levels of noise (this is a interest of future work).

Although we assume spatially coherent illumination in our simulations above, there is no hard constraint on this requirement. Our algorithm can be modified to work in settings of partial coherence by accounting for the partial coherence in the deconvolution step as per [63], and by changing the noisy matrix completion technique in Fig. 4-6's dotted box to keep the highest k singular value instead of just one (if a maximum rank of the mutual intensity is known) or changing the matrix completion technique altogether (e.g., to singular value thresholding [27] if a maximum rank of the mutual intensity matrix is not known).

Additionally, although we present low-rank matrix completion as a technique for denoising Wigner Distribution Deconvolution in a conventional ptychography-like setup, our algorithm is extendible to a Fourier ptychography-like setup (this is the subject of current work). The Fourier ptychography and ptychography data matrices are transposes of one another, and by transposing each step of WDD to its Fourier dual, one can obtain an estimate of the object's mutual intensity after which LRMR/LRMC could be used to denoise that estimate and recover the object's amplitude and phase.

In its original form, WDD is intrinsically limited in resolution by the probe step size s_i (usually significantly larger than detector pixel size), making the technique inferior to ptychography (and Fourier ptychography) in terms of resultant space-

bandwidth product. However, it is possible that further modifications to the algorithm (e.g., using interpolated estimates for measurements in between actual measurements taken at large step size to form $I(u, s)$ and denoising error generated from this step later) may enable WDD to perform competitively against ptychography not just for phase retrieval but also in terms of the space-bandwidth product of the reconstructed image. Furthermore, simply having a noise-robust low-resolution initial estimate of an object's amplitude and phase may signficantly benefit the convergence rate and accuracy of some ptychographic reconstruction algorithms such as Wirtinger Flow[79].

Accurate results in Ptychography and Fourier Ptychography are often dependent on low-noise/high-photon count setups, and these setups often require high-bit depth sensors in well-controlled (e.g., cooled) settings. However, in order for ptychographic imaging to gain traction outside of research settings (e.g., one of our goals is to bring ptychographic imaging into use for clinical pathology), reconstruction techniques must work with lower-end sensors and in situations with high noise. Denoised WDD may provide a step towards this objective as our results suggest that a class of phase retrieval techniques relying on regularized direct inversion of ptychographic datasets via WDD (instead of iterative reconstruction techniques) can provide accurate quantitative phase information in the presence of high levels of noise.

Chapter 5

Lensless Computational Imaging via Deep Leaning

This chapter is comprised of three sections. The first section reviews three lensless imaging techniques: in-line digital holography, Gerchberg-Saxton-Fienup (GSF) iterative phase retrieval, and transport of intensity equation (TIE) imaging. At the end of this first section, the author presents a novel method for phase retrieval via deep learning given defocused image inputs (as in TIE imaging). The second section of this chapter expands upon this initial idea and experimentally validates end-to-end lensless computational imaging through deep learning. Contents of this section were adapted from the published paper, "Lensless computational imaging through deep learning [122]," for which the author of this thesis contributed the initial code, experimental setup and deep learning model. The final section of this chapter touches briefly upon extensions of the work presented here.

5.1 Computational Imaging Techniques

Computational imaging utilizes software to augment or replace the role of traditional optical elements in imaging systems. By designing imaging hardware and software in concert, imaging systems can be made simpler/more robust and have capabilities beyond that of traditional imaging systems. Examples of popular computational

imaging setups include: digital holography, iterative phase retrieval techniques such as the Gerchberg-Saxton Algorithm, and transport of intensity equation (TIE) imaging.

5.1.1 In-line Digital Holography

In-line digital holography seeks to recover a complex index distribution of three-dimensional(3D) objects from measured wavefront information. The in-line digital holography setup is simple – a scattering object (or set of objects) is illuminated with a spatially coherent, monochromatic plane wave, and a Gabor (in-line) hologram is formed by the interference pattern of the original wave, E_r, with the wavefront generated by the illuminated objects with scattering potential $\eta(x', y', z')$. Since cameras can only record intensity at optical frequencies, a 2-D detector array records the irradiance incident on the camera:

$$I(x, y) = |E_r + E_s|^2 = |E_r|^2 + |E_s(x, y)|^2 + E_r^* E(x, y) + E_r E^*(x, y)$$

Inversion from this single 2D intensity measurements to the originating 3D scattering potential is an underdetermined problem, but techniques have developed to reconstruct an estimate of the underlying 3D scattering potential based on focus metrics [145, 146] and sparsity constraints [147, 146]. Figure 5-1 reconstruction results of shows one such algorithm that the author has implemented to image bubbles in a fish tank.

5.1.2 Iterative Phase Retrieval: Gerchberg-Saxton Fienup Algorithm

The Gerchberg-Saxton (GS) algorithm is an iterative algorithm for phase retrieval which alternates between two (or more) planes related by a propagating function (e.g., a Fourier Transform or Fresnel Propagation), with only intensity observable/known at these plane (i.e., phase is unknown).

The GS algorithms begins with a random initial guess for the phase at a chosen

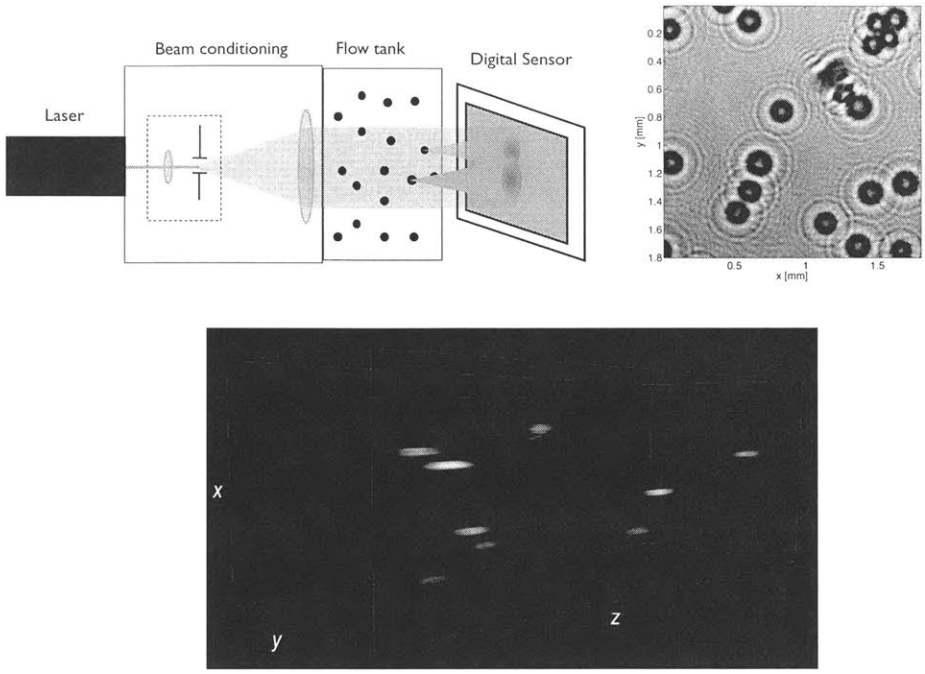

Figure 5-1: "Compressive holographic inversion of particle scattering." (top left) In-line DH experimental setup. A 3D scattering potential of bubbles in water was reconstructed (bottom) from a single 2D snapshot (top right) using compressive holography.

plane (let's call it the "input plane" although it could physically be the "output plane"). This plane's complex field estimate (amplitude + phase) is then forward-propagated to obtain an estimate of the true field at the output plane. This estimated output plane intensity is then compared to the true or desired output intensity. If the error is large (such that the system has not reached an acceptable solution yet), then the estimated output plane amplitude is replaced with the true amplitude of the output plane while its phase is kept the same. This new estimated complex output field is then back-propagated to obtain an estimate of the true field at the input plane. The newly estimated input plane intensity is then compared to the true or desired input intensity. If the error is large, then the newly estimated input plane amplitude is replaced with the true amplitude of the input plane while its phase is kept the same. This new input plane estimate is forward-propagated and the entire process above repeats in a "ping-pong"-like fashion until convergence.

The GS method of "alternating projections" has the significant drawback of often

converging very slowly to poor local minima, but results can be significantly improved by applying other input constraints such as positivity, finite-support, or smoothness. Additionally, a hybrid input-output algorithm as proposed by Fienup [148, 149] has been shown to signifcantly speed up GSF algorithm convergence.

5.1.3 Transport of Intensity Imaging

Transport of Intensity Equation (TIE) imaging addresses the phase retrieval problem by taking multiple intensity measurements at multiple planes in space. The general idea behind TIE imaging is that although a single intensity measurement cannot provide phase information, if one knows the intensity distribution of light as it propagates over a volume, then he or she can directly recover the phase of light based on principles of energy transport.

The TIE is given as [123]:

$$\frac{2\pi}{\lambda}\frac{\partial I}{\partial z} = \nabla_{\perp} \cdot I \nabla_{\perp} \phi$$

where I is the intensity measured in a plane perpendicular to the axis of propagation, z, and the $\perp$ subscripts indicate the operator acting in a plane perpendicular to the optical axis. λ is the spectrally-weighted mean wavelength of illumination [123], and ϕ is the phase. At its root, the TIE describes a conservational principle: the perpendicular gradient of the phase of coherent light represent the flux of light in/outwards in the x and y directions. The divergence of that quantity denotes the "source or sink" nature which arises due to the intensity change from z-plane to z-plane.

5.1.4 Deep Learning Transport of Intensity Imaging

Here, we propose and demonstrate in simulation an algorithm using deep learning for phase retrieval using a focal stack of images similar to those that would be obtained for TIE imaging. At the time of its creation, this demonstration was, to the best of our knowledge, the first demonstration of deep learning's ability to solve the phase retrieval problem.

To generate our training set for learning TIE-based phase retrieval, we used 100,000 images from the ImageNet [120] database. These images were then split into two equal sets of 50,000 – one to use as training set object amplitudes and the other to use as training set object phases. Images from these two sets were then randomly paired to generate 50,000 mixed amplitude-phase images. A test set of 10,000 amplitude-phase images were generated in a similar fashion.

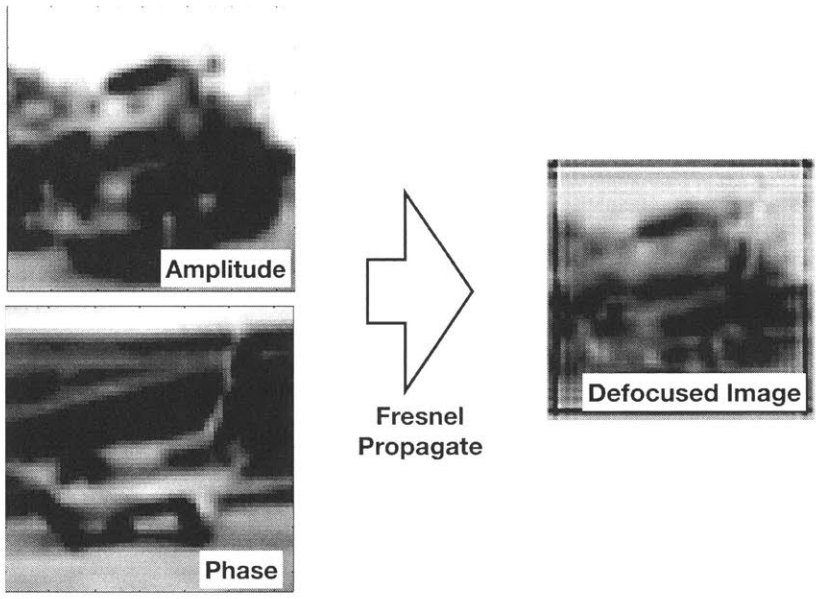

Figure 5-2: Deep Learning Transport of Intensity Imaging

These mixed amplitude-phase images were Fresnel propagated to generate defocused "recorded" image intensities (Figure 5-2), and these defocused images, along with the in-focus intensity images and a "dIdz" first-order derivative estimate (obtained by differencing the defocused and in-focus images) were fed as inputs into a simple convolutional neural network (five convolution + ReLU blocks) to estimate object phase at the in-focus plane. Results on the test set after training 20 epochs (SGD, lr = 0.02, batchsize=16) are shown in Figure 5-3. Training for this demo was performed using MatConvNet (later experiments use TensorFlow).

The results, although far from perfect, demonstrate that deep neural networks have the ability to reconstruct phase from a set of in-focus and defocused images. This demonstration forms the basis/groundwork for efforts in the following section.

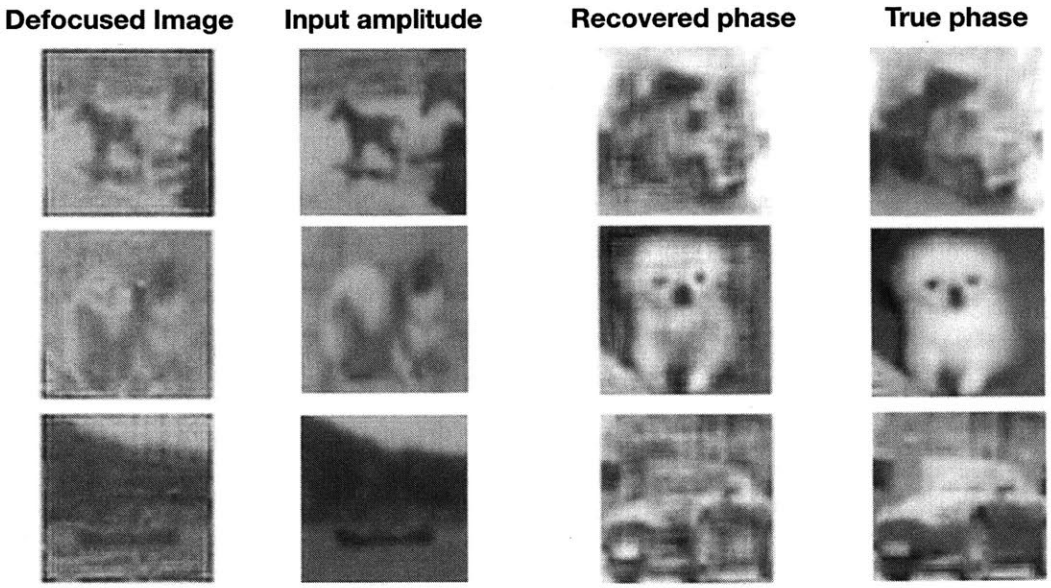

Figure 5-3: Deep Learning Transport of Intensity Imaging

5.2 Lensless computational imaging through deep learning

Deep learning has been proven to yield reliably generalizable solutions to numerous classification and decision tasks. Here, we demonstrate for the first time to our knowledge that deep neural networks (DNNs) can be trained to solve end-to-end inverse problems in computational imaging. We experimentally built and tested a lensless imaging system where a DNN was trained to recover phase objects given their propagated intensity diffraction patterns

5.2.1 Introduction

Neural network training can be thought of as generic function approximation: given a training set (*i.e.*, examples of matched input and output data obtained from a hitherto-unknown model), a neural network attempts to generate a computational architecture that accurately maps all inputs in a test set (distinct from the training set) to their corresponding outputs. The basic idea dates back to the 1950's [113]; recently, Deep Neural Networks (DNNs), called so because they are structured as

multi-layered with a large number of layers, have found extensive use. To generalize well, DNNs exploit several innovations in connectivity, for example convolutions [111, 109, 125, 98] for regularization and pruning in image recognition and classification tasks; nonlinearities, such as non-differentiable piecewise linear units [94] as opposed to the older sigmoidal functions that were differentiable but also prone to stagnation [96]; and algorithms, such as more efficient back propagation (backprop) [119, 112]. In turn, these have been applied in diverse domains: playing complex games on Atari 2600 [114] and Go [121]; object generation [92]; object detection [110]; and image processing: colorization [89], deblurring [90, 130, 124], and in-painting [129]. In this paper, we propose that deep neural networks may "learn" to approximate solutions to inverse problems in computational imaging.

A general computational imaging system consists of a physical part and a computational part. In the physical part, light propagates through one or more objects of interest as well as optical elements such as lenses, prisms, etc. finally producing a raw intensity image on a digital camera. The raw intensity image is then computationally processed to yield object attributes, e.g. a spatial map of light attenuation and/or phase delay through the object—what we traditionally call "image" and "quantitative phase image," respectively. The computational part of the system is then said to have produced a solution to the inverse problem.

The specific computational imaging scenario chosen for this section is quantitative phase retrieval. Let $\psi(x, y; z = 0) = \exp\left[i\phi(x, y)\right]$ represent the optical field purely phase-modulated by the unknown object $\phi(x, y)$ as function of the lateral Cartesian coordinates x, y at the origin $z = 0$ of the optical axis. The measurement after propagation by distance z is the raw intensity image $I(x, y) = |\psi(x, y; z)|^2 \equiv H\phi(x, y)$, where H denotes the forward operator relating the phase $\phi(x, y)$ at the origin to the raw intensity image $I(x, y)$ at distance z. The problem of phase retrieval, then, is to formulate an inverse transform H^{inv}, generally nonlinear, such that

$$\hat{\phi}(x, y) = H^{\text{inv}} I(x, y) \tag{5.1}$$

represents an acceptable estimate of the phase object. The Gerchberg-Saxton-Fienup (GSF) class of algorithms [95, 93] is a well-known and highly successful approach where instead of an explicit H^{inv}, the inverse $\hat{\phi}(x,y)$ is estimated iteratively from a single raw intensity image. Two other notable approaches are: digital holography (DH) [97], where an additional reference beam $\exp[ikx]$ is assumed to be present so that instead an interferometric measurement $I_{\text{holo}} = |\exp[ikx] + \psi(x,y;z)|^2$ is available; and transport of intensity (TIE) [126, 123], utilizing an "axial stack" of one or more additional raw intensity images $I_1(x,y) = |\psi(x,y;z+\delta z_1)|^2, I_2(x,y) = |\psi(x,y;z+\delta z_2)|^2, \ldots$, where $\delta z_1, \delta z_2, \ldots$ are axial displacements. On-axis DH ($k = 0$) requires the object to be spatially sparse so that the incoming light remains mostly unscattered because it must serve as reference beam for the digital hologram; whereas off-axis DH ($k \neq 0$) requires the object to be low-bandwidth so that the space-bandwidth product limitation of the camera measuring I_{holo} isn't violated. On the other hand, TIE requires sparse object gradients to avoid singularities as result of propagation.

Another approach to obtain H^{inv} is by solving an optimization problem of the form

$$\hat{\phi}(x,y) = \text{argmin}_{\phi} \|H\phi - I\|^2 + \alpha\Phi(\phi). \tag{5.2}$$

Here, H is the same forward operator, I is the measurement (including noise), Φ is the regularizer expressing prior information about the object, and α is the regularization parameter that controls the relative strength of the two terms in the optimization functional. This computational approach, central to many inverse problems [87], is traced back to Tikhonov [127] and Wiener [128]. For (5.2) to work well, in addition to the obvious concerns of convexity and convergence speed, H and Φ need to be precisely known or at least discoverable. For Φ, expressions promoting sparsity have been often proven to be effective [77, 88, 91]; on the other hand, the damage due to incomplete or erroneous knowledge in H, for example due to insufficient accuracy in the approximations or due to experimental errors such as misalignments, is more difficult to control [104].

The hypothesis that we set out to test in this paper is whether a DNN can be trained to act as the inverse operator H^{inv} in (5.1). In addition to the well-established adaptability of DNNs as effective approximators to general-purpose nonlinear operators, our approach is attractive for several reasons: the DNN, once trained, should recover every possible object within a hopefully large enough class, unlike GSF where H^{inv} is obtained separately and iteratively for each new object presented. Alternatively, since during DNN training iteration takes place over several objects simultaneously, the learning approach is expected to be a more robust generalization of GSF. Moreover, the DNN approach is not subject to any of the sparsity requirements of either DH or TIE. Indeed, our experiment, described in detail in Section 5.2.2, was explicitly designed to violate the DH and TIE assumptions. Lastly, compared to optimization-based approaches, the DNN input-output relationship (which is in general too complicated to be written in explicit form) replaces the form (5.2). The user is therefore not required to specify the forward operator H and the priors Φ or decide about the magnitude of the regularization parameter α; instead, to the degree that our results show successful inversion, they suggest that the DNN learns H, Φ, and α implicitly through training. The downside of the DNN approach is that a sufficiently large database of known pairs of phase objects and their corresponding raw intensity images must be available for training and testing; we will return to this issue shortly.

Neural networks have been used in a variety of imaging and image restoration tasks. For example, Hopfield's associative memory network [100] was capable of retrieving entire faces from partially obscured inputs, and was implemented in an all-optical architecture [103]. Recently, Horisaki et al. [101] used support-vector machines, an older form of bi-layer neural network with nonlinear discriminant functions, also to recover face images through a scattering medium; the specific scatterer was learnt by the neural network *in lieu* of the more common approach of having the scatterer characterized through a transmission matrix [115, 116]. Modern deep networks and training schemes have been used to model (learn) the scatterer's forward operator using an architecture designed to emulate the layered structure of the scatterer itself [107]. More recently, DNNs have also been used for classification of cells from raw

intensity images captured in a holographic arrangement [106]; and improvement in image quality by first forming images with traditional means and then using a DNN trained from object-image pairs to sharpen the visual image quality. DNN-enabled image restoration has been carried in a variety of forward operators: tomography [105], digital holography [118] and standard microscopy [117]. Our approach is, instead, truly end-to-end: the DNN implements the H^{inv} operator from (5.1) in its entirety; the first such end to end attempt, to our knowledge, is an earlier, more primitive effort by our own group with binary phase objects [122]. Moreover, because the phase objects in both our earlier and current efforts are generated by a precisely calibrated spatial light modulator, we are able to quantitatively evaluate the fidelity of our DNN-generated estimates $\hat{\phi}(x, y)$ to the true objects $\phi(x, y)$.

Neural network approaches often come under criticism because the quality of training depends on the quality of the examples given to the network during the training phase. For instance, if the inputs used to train a network are not diverse enough, then the DNN will learn priors of the input images instead of generalized rules for phase retrieval from intensity. This was the case in [101], where an SVM trained using images of faces could adequately reconstruct faces, but when given the task of reconstructing images of natural objects such as a pair of scissors, the trained SVM still returned an output that resembled a human face.

For our specific problem, an ideal training set would encompass all possible phase objects. Unfortunately, phase objects, generally speaking, constitute a rather large class and it would be unrealistic to attempt to train a network sampling from across all possible objects from this large class. Instead, we synthesize phase objects in the form of natural images derived from the ImageNet [120] database because it is readily available and widely used in the study of various machine learning problems. For comparison, we also trained a separate network using a narrower class of (facial) images from the Faces-LFW [102] database.

As expected, our network did well when presented with unknown phase objects in the form of faces or natural images that it had been trained to. Notably, the network also performed well when presented with objects outside of its training class – the

DNN trained using images of faces was able to reconstruct images of natural objects, and the DNN trained using images of natural objects was able to reconstruct images of faces. Additionally, both DNNs were able to reconstruct completely distinct images including: handwritten digits, characters from different languages (Arabic, Mandarin, English), and images from a disjoint natural image dataset. Both trained networks yielded accurate results even when the object-to-sensor distance(s) in the training set slightly differed from that of the testing set, suggesting that the networks are not merely pattern-matching but instead have actually learned a generalizable model approximating the inverse operator H^{inv}.

The details of our experiment, including the physical system and the computational training and testing results, are described in Section 5.2.2. The neural network itself is analyzed in Section 5.2.3, and concluding thoughts are in Section 5.2.4.

5.2.2 Experiment

Our experimental arrangement is as shown in Figure 5-4. Light from a He-Ne laser source (Thorlabs, HNL210L, 632.8nm) first transmits through a spatial filter, which consists of a microscope objective (Newport, M-60X, 0.85NA) and a pinhole aperture ($D = 5\mu$m), to remove spatial noise. After being collimated by the lens ($f = 150$mm), the light is reflected by a mirror and then passes through a linear polarizer, followed by a beam splitter. A spatial light modulator (Holoeye, LC-R 720, reflective) is placed normally incident to the transmitted light and acts as a pixel-wise phase object. The SLM pixel size is $20 \times 20\mu\text{m}^2$ and number of pixels is 1280×768, out of which the central 512×512 portion only is used in the experiments. The SLM-modulated light is then reflected by the beam splitter and passes through a linear polarization analyzer, before being collected by a CMOS camera (Basler, A504k) placed at a distance z. The CMOS camera pixel size is $12 \times 12\mu\text{m}^2$ and number of pixels is 1280×1024, which is cropped to a 1024×1024 square used for processing. The total used linear camera size of ≈ 12.3mm is slightly larger than the active SLM size of ≈ 10.2mm to accommodate expansion of the propagating beam due to diffraction. Images recorded by the CMOS camera are then processed on an Intel i7 CPU, with neural network

computations performed on a GTX1080 graphics card (NVIDIA).

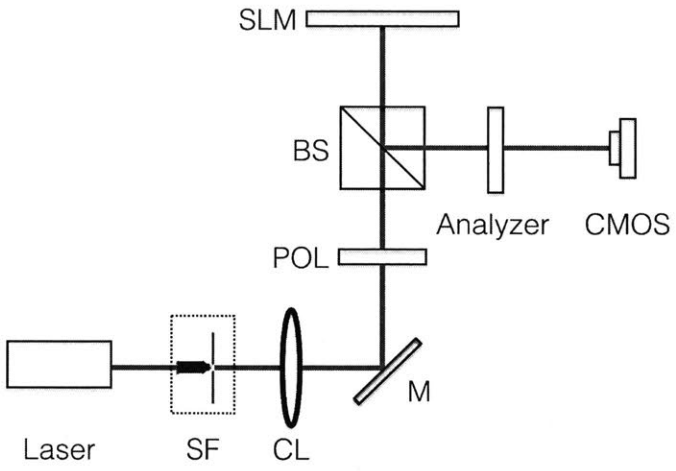

Figure 5-4: Experimental arrangement. SF: spatial filter; CL: collimating lens; M: mirror; POL: linear polarizer; BS: beam splitter; SLM: spatial light modulator.

According to its user manual, the LC-R 720 SLM can realize (approximate) pure-phase modulation if we modulate the light polarization properly. Specifically, for He-Ne laser light, if we set the incident beam to be linearly polarized at 45° with respect to the vertical direction and also set the linear polarization analyzer to be oriented at 340° with respect to the vertical direction, then the amplitude modulation of the SLM will become almost independent of the assigned (8-bit gray-level) input. In this arrangement, the phase modulation of the SLM follows a monotonic, almost-linear relationship with the assigned pixel value (with maximum phase depth: $\sim 1\pi$). We experimentally evaluated the correspondence between 8-bit grayscale input images projected onto the SLM and phase values in the range $[0, -\pi]$ (see supplement). In this paper, we approximate our SLM as a pure-phase object and computationally recover the phase using a neural network.

The CMOS detector was placed after a free-space propagation distance z. Distinct DNNs were trained from recorded diffraction patterns at three distances $z_1 =$ 37.5cm$\pm$2mm (NA=0.0164$\pm 9\times 10^{-4}$, Fresnel number $\mathcal{F}$=159$\pm$1), $z_2 =$ 67.5cm$\pm$2mm (NA=0.0091 $\pm$ 1 $\times$ 10^{-4}, Fresnel number $\mathcal{F}$=88.3 $\pm$ 0.3) and $z_3 =$ 97.5cm $\pm$ 2mm (NA=0.0063 $\pm$ 1 $\times$ 10^{-4}, Fresnel number $\mathcal{F}$=61.2 $\pm$ 0.1). Our experiment consisted of two phases: training and testing. During the training phase, we modulated the

phase SLM according to samples randomly selected from the Faces-LFW or ImageNet database. We resized and padded selected images before displaying them on our SLM. Two examples of inputs, as they are sent to the SLM, and their corresponding raw intensity images (diffraction patterns) as captured on the CMOS are shown in Figure 5-5. Our training set consisted of 10,000 such faces/images - diffraction pattern pairs. The raw intensity images from all these training examples were used to train the weights in our DNN. We used a Zaber A-LST1000D stage with repeatability $2.5\mu m$ to translate the camera in order to analyze the robustness of the learnt network to axial perturbations. The positional accuracy of $\pm$2mm reported earlier on the training distances z_1, z_2, z_3 is derived from our error in manually establishing the absolute distance between SLM and camera; whereas the stage repeatability determines the error in camera displacement relative to these initial training positions, respectively, for each axial robustness test in Section 5.2.3.

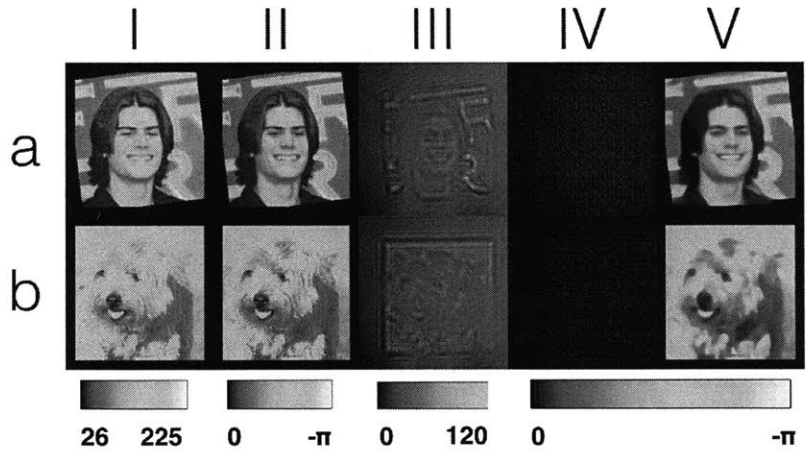

Figure 5-5: DNN training. Rows (a) and (b) denote the networks trained on Faces-LFW and ImageNet dataset, respectively. (i) randomly selected example drawn from the database; (ii) calibrated phase image of the drawn sample; (iii) diffraction pattern generated on the CMOS by the same sample; (iv) DNN output before training (*i.e.* with randomly initialized weights); (v) DNN output after training.

Our DNN uses a convolutional residual neural network (ResNet) architecture. In a convolutional neural network (CNN), inputs are passed from nodes of each layer to the next, with adjacent layers connected by convolution. Convolutional ResNets extend CNNs by adding short term memory to each layer of the network. The intuition

behind ResNets is that one should only add a new layer if one stands to gain by adding that layer. ResNets ensure that the $(N+1)$th layer learns something new about the network by also providing the original input to the output of the $(N+1)$th layer and performing calculations on the residual of the two. This forces the new layer to learn something different from what the input has already encoded/learned [98].

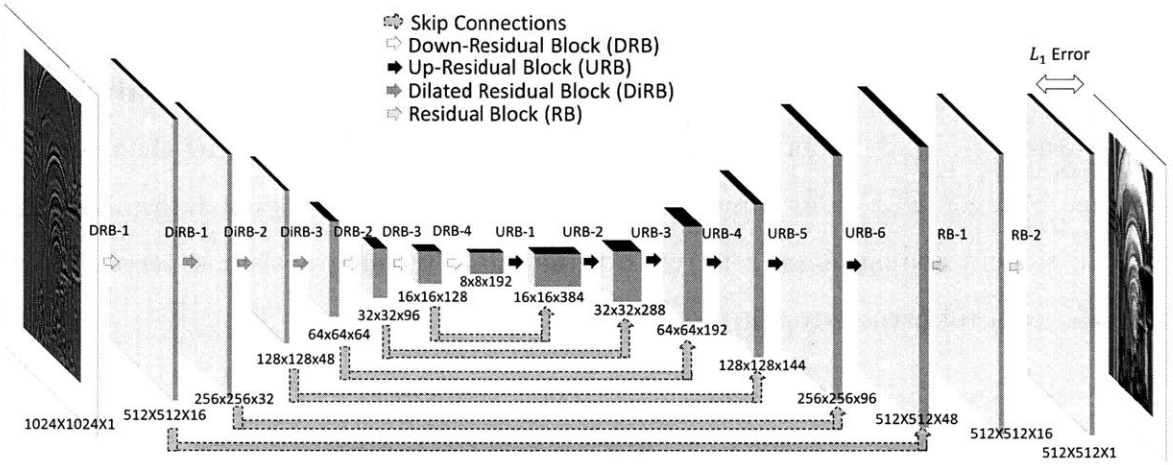

Figure 5-6: Detailed schematic of our DNN architecture, indicating the number of layers, nodes in each layer, etc.

A diagram of our specific DNN architecture is shown in Fig. 5-6. The input layer is the image captured by the CMOS camera. It is then successively decimated by 7 residual blocks of convolution + downsampling followed by 6 residual blocks of deconvolution + upsampling, and finally 2 standard residual blocks. Some of the residual blocks comprise dilated convolutions so as to increase the receptive field of the convolution filters, and hence, aggregate diffraction effects over multiple scales [131]. We use skip connections to pass high frequency information learnt in the initial layers down the network towards the output reconstruction, and have two standard residual blocks at the end of the network to finetune the reconstruction. At the very last layer of our DNN, the values represent an estimate of our phase object. The connection weights are trained using backpropagation (not to be confused with optical backpropagation) on the L_1 distance between the network output and the

nominal appearance of the training samples represented as:

$$\min \frac{1}{wh} \parallel Y - G \parallel_1 . \tag{5.3}$$

Here, w, h are the width and the height of the output, Y is the output of the last layer, and G is the ground truth phase value, limited in the range $[0, -\pi]$. Additional details about the training of the DNN are provided in the Supplement, Section 2.

We collected data from six separate experiment runs using training inputs from Faces-LFW or ImageNet and object-to-sensor distances of z_1, z_2 , or z_3. These data were used to train six separate DNNs for evaluation.

Fig. 5-5(iv) shows a sample DNN's output at the beginning of its training phase (*i.e.* with randomly initialized weights), and Fig. 5-5(v) shows the network output after training, for the same example object-raw image pairs. Training each network took $\approx$ 20 hours using Tensorflow on our GPU. We provide analysis of the trained DNN in Section 5.2.3.

Our testing phase consisted of: (1) sampling disjoint examples from the same database (either Faces-LFW or ImageNet) and other databases such as MNIST, CIFAR, Faces-ATT etc., (2) using these test examples to modulate the SLM and produce raw intensity images on the camera, (3) passing these intensity images as inputs to our trained DNN, and (4) comparing the output to ground truth.

5.2.3 Results and Network analysis

The standard method of characterizing neural network training is by plotting the progression of training and test error across training epochs (iterations in the back-propagation algorithm over all examples). These curves are shown in Figure 5-7 for our network trained using the ImageNet database and tested using images from: (a) Faces-LFW (b) a disjoint ImageNet set, (c) images from an English/Chinese/Arabic characters database, (d) the MNIST handwritten digit database, (e) Faces-ATT [86], (f) CIFAR [108], (g) a constant-value "Null" image. Our ImageNet learning curves in Figure 5-7(d) show convergence to low value after $\sim$10 epochs, indicating that our

network has not overfit to our training dataset. We plot bar graphs for the mean absolute error (MAE) over test examples in the 7 different datasets for each of the 3 object-to-sensor distances in Figure 5-7. Lower MAE was reported for test images with large patches of constant value (characters, digits, Null) as their sparse diffraction patterns were easier for our DNN to invert. Notably, both our bar graphs and learning curves show low test error for the non-trained images, suggesting that our network generalizes well across different domains.

This is an important point and worth emphasizing: despite the fact that our network was trained exclusively on images from the ImageNet database – i.e., images of planes, trains, cars, frogs, artichokes, etc., it is still able to accurately reconstruct images of completely different classes (e.g., faces, handwritten digits, and characters from different languages). This strongly suggests that our network has learned a model of the underlying physics of the imaging system or at the very least a generalizable mapping of low-level textures between our output diffraction patterns and input images.

A more pronounced qualitative example demonstrating this is shown in the columns (iv) (vii) and (x) of Figure 5-8. Here, we trained our network using images exclusively from the Faces-LFW database. Despite this limited training set, the learned network was able to accurately reconstruct images from the ImageNet, handwritten digits, and characters datasets. This is in contrast to results shown in [101], where an SVM trained on images of faces was able to accurately reconstruct images of faces but not other classes of objects.

We tested the robustness of our network to rotation and lateral displacement in the presented test phase objects, as well as axial displacement of the CMOS camera for each DNN relative to the DNN's trained axial locations z_1, z_2, z_3, respectively. Quantitative results of these perturbations are shown in Figures 5-9, 5-10, 5-11, and qualitative results for the networks trained at distance z_1 are shown in Figures A-7, 5-13 and 5-14. Qualitative results for the other two distances are in the supplement. The results show that our trained network is robust to moderate perturbations in sensor displacement and is somewhat shift and rotation invariant. As expected, the

reconstruction fails when the displacement is significantly greater (Figure A-9).

To get a sense of what the network has learned, we examined its maximally-activated patterns (MAPs), i.e., what types of inputs would maximize network filter response (gradient descent on the input with average filter response as loss function [132]). Our results are shown in Figure 5-16 together with the results of analogous analysis of a deblurring network of similar architecture as well as an ImageNet classification DNN. Compared with MAPs of ImageNet and a deblurring network, the MAPs of our phase-retrieval network show two primary differences: First, compared to ImageNet, we observe much finer textures; this is because ImageNet is meant to do classification, a task that requires high-level features to emerge out of learning; whereas our phase network is performing a form of regression. Secondly, compared with the deblurring network, we observe somewhat finer textures, especially at the shallower layers (although in both cases low-level textures are present). Our interpretation is that both phase retrieval and deblurring require local operations but of different nature: deblurring converts intensity gradients into sharp edges, whereas phase retrieval converts diffraction rings into phase gradients. The difference between the two cases is not easily discernible by simple visual inspection of the MAPs. The point is that phase retrieval and deblurring share this common locality feature while acting differently for the sake of their respective functions; and both are radically different than classification networks, such as ImageNet.

5.2.4 Conclusions and discussion

The architecture presented here was deliberately well controlled, with an SLM creating the phase object inputs to the neural network for both training and testing. This allowed us to quantitatively and precisely analyze the behavior of the learning process. Application-specific training, e.g. replacing the SLM with physical phase objects for more practical applications, we judged beyond the scope of the present work. Other obvious and useful extensions would be to include optics, e.g. a microscope objective for microscopic imaging in the same mode; and to attempt to reconstruct complex objects, *i.e.* imparting both attenuation and phase delay to the incident light. The

significant anticipated benefit in the latter case is that it would be unnecessary to characterize the optics for the formulation of the forward operator—the neural network should "learn" this automatically as well. We intend to undertake such studies in future work.

5.3 Future Work

Several natural extensions exist for the work presented in this chapter, many of which are the subject of current study. One such extension is the application of our imaging system for microscopic imaging [144, 142]. In such a setup, one may train the neural network to learn generalized phase retrieval (i.e., over a broad class of objects) using an SLM placed in a microscope setup. Once trained, the deep learning microscope may be used without the SLM to image microscopic objects of interest. Alternatively, one may also train the network using domain specific images (e.g., of a particular type of cell or biological tissue) [143] if it is known that the final imaging platform will only be used to image a particular class of objects.

Another extension of the work presented in this chapter is utilization of our imaging system and deep learning pipeline to image through turbid media [141]. From a neural network input-output perspective, the problems are very similar: the desired network output is a ground truth "system input" 2D image/signal, and network inputs are filtered versions of this signal (e.g., by Fresnel propagation or passage through scattering media). This input-output relationship also has similarities to DNNs used for image deblurring or denoising. However, DNNs used for imaging (vs. image processing) have inputs that cannot be perfectly registered with outputs (due to imperfect experimental setups), and for the case of imaging through turbid media, significantly larger receptive fields are necessary in order to access information from light scattered at high angles. Deep learning phase retrieval from focal stack images or phase-space representations may help address this problem and is also an area of active interest.

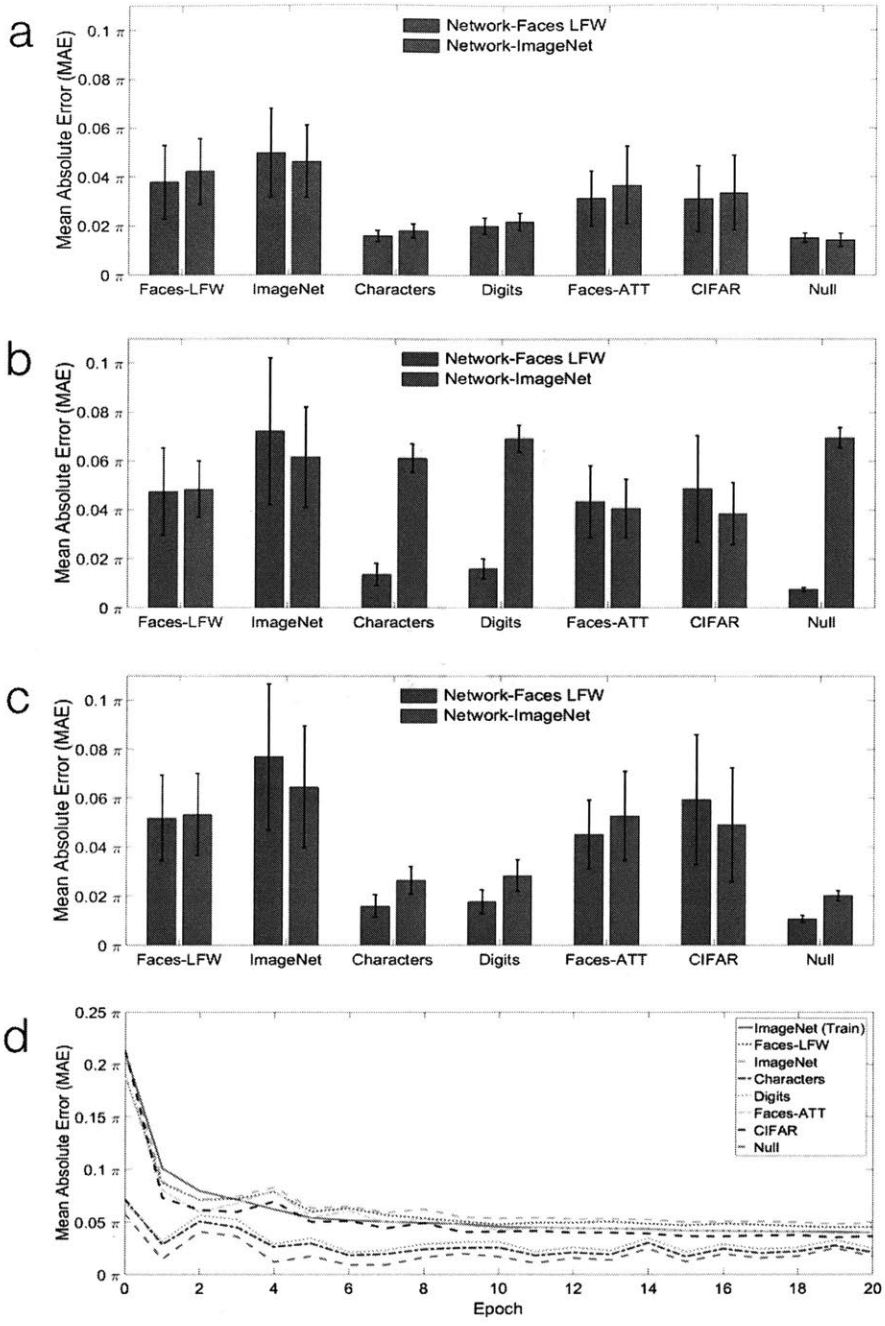

Figure 5-7: Quantitative analysis of our trained deep neural networks for three object-to-sensor distances (a) z_1, (b) z_2, and (c) z_3 for the DNNs trained on Faces-LFW (blue) and ImageNet (red) on seven datasets. (d) The training and testing error curves for network trained on ImageNet at distance z_3 over 20 epochs.

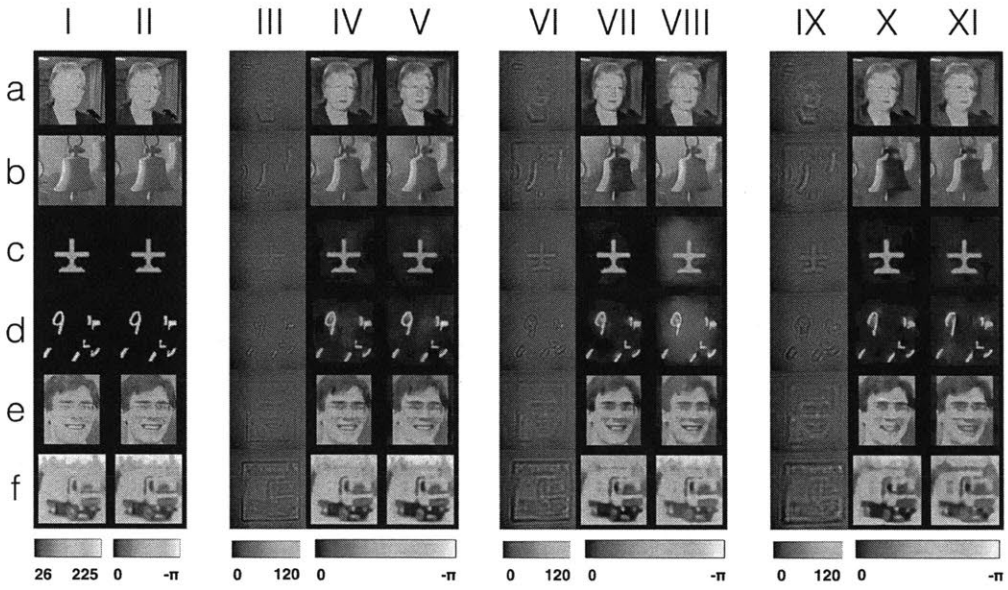

Figure 5-8: Qualitative analysis of our trained deep neural networks for combinations of object-to-sensor distances z and training datasets. (i) Ground truth pixel value inputs to the SLM. (ii) Corresponding phase images calibrated by SLM response curve. (iii) Raw intensity images captured by CMOS detector at distance z_1. (iv) DNN reconstruction from raw images when trained using Faces-LFW [102] dataset. (v) DNN reconstruction when trained used ImageNet [120] dataset. Columns (vi-viii) and (ix-xi) follow the same sequence as (iii-v) but in these sets the CMOS is placed at a distance of z_2 and z_3, respectively. Rows (a-f) correspond to the dataset from which the test image is drawn: (a) Faces-LFW, (b) ImageNet, (c) Characters, (d) MNIST Digits, (e) Faces-ATT [86], or (f) CIFAR [108], respectively.

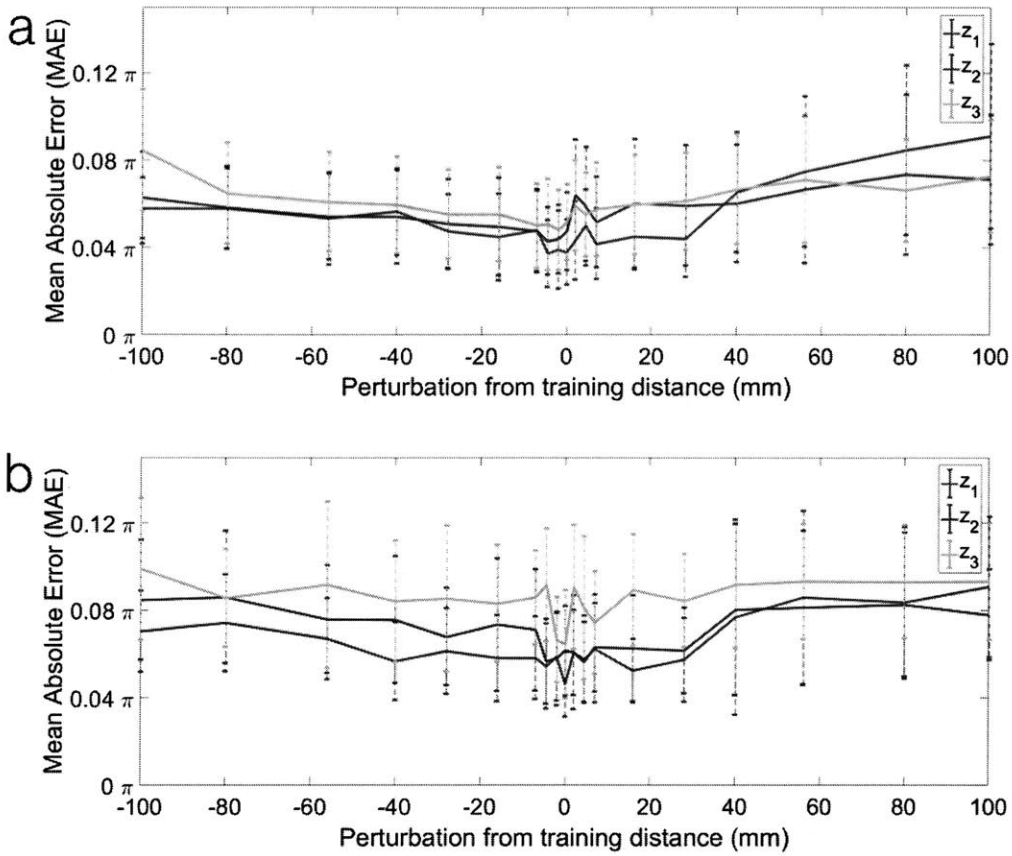

Figure 5-9: Quantitative analysis of the sensitivity of the trained deep convolutional neural network to the object-to-sensor distance. The network was trained on (a) Faces-LFW database and (b) ImageNet and tested on disjoint Faces-LFW and ImageNet sets, respectively. The nominal depths of field for the three corresponding training distances z_1, z_2, z_3, respectively, are: $(\text{DOF})_1 = 1.18 \pm 0.1\text{mm}$, $(\text{DOF})_2 = 3.82 \pm 0.2\text{mm}$, and $(\text{DOF})_3 = 7.97 \pm 0.3\text{mm}$.

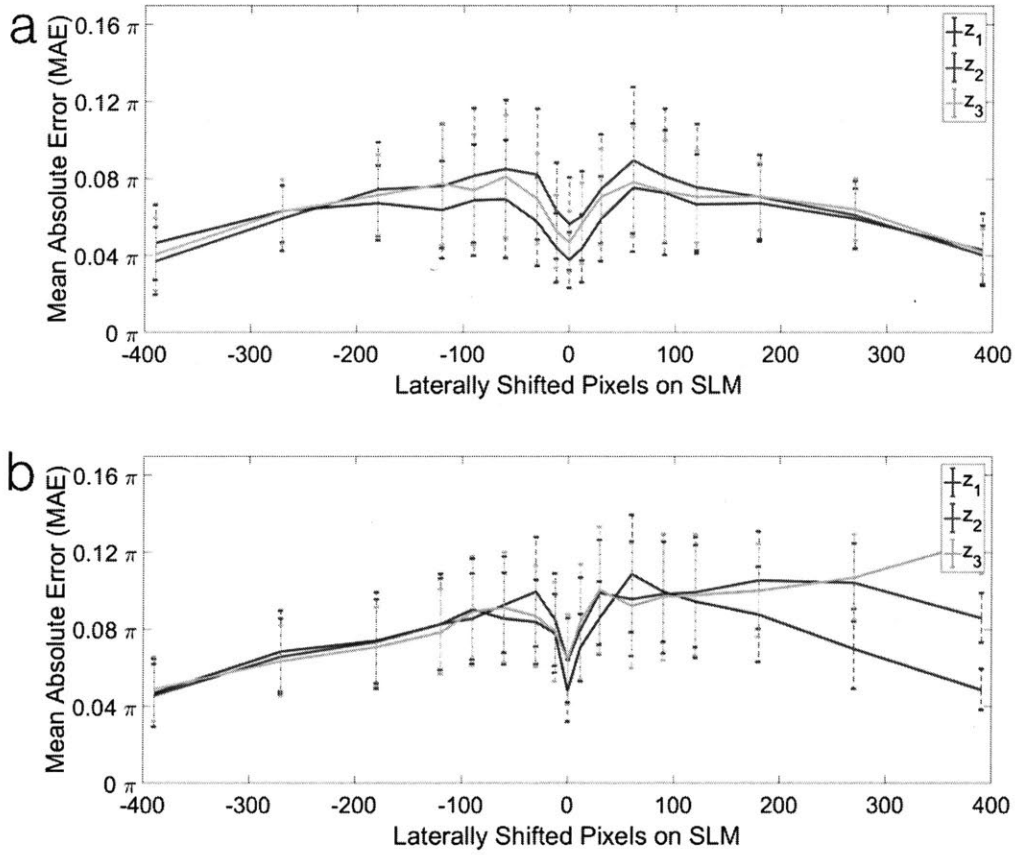

Figure 5-10: Quantitative analysis of the sensitivity of the trained deep convolutional neural network to laterally shifted images on the SLM. The network was trained on (a) Faces-LFW database, (b) ImageNet and tested on disjoint Faces-LFW and ImageNet sets, respectively.

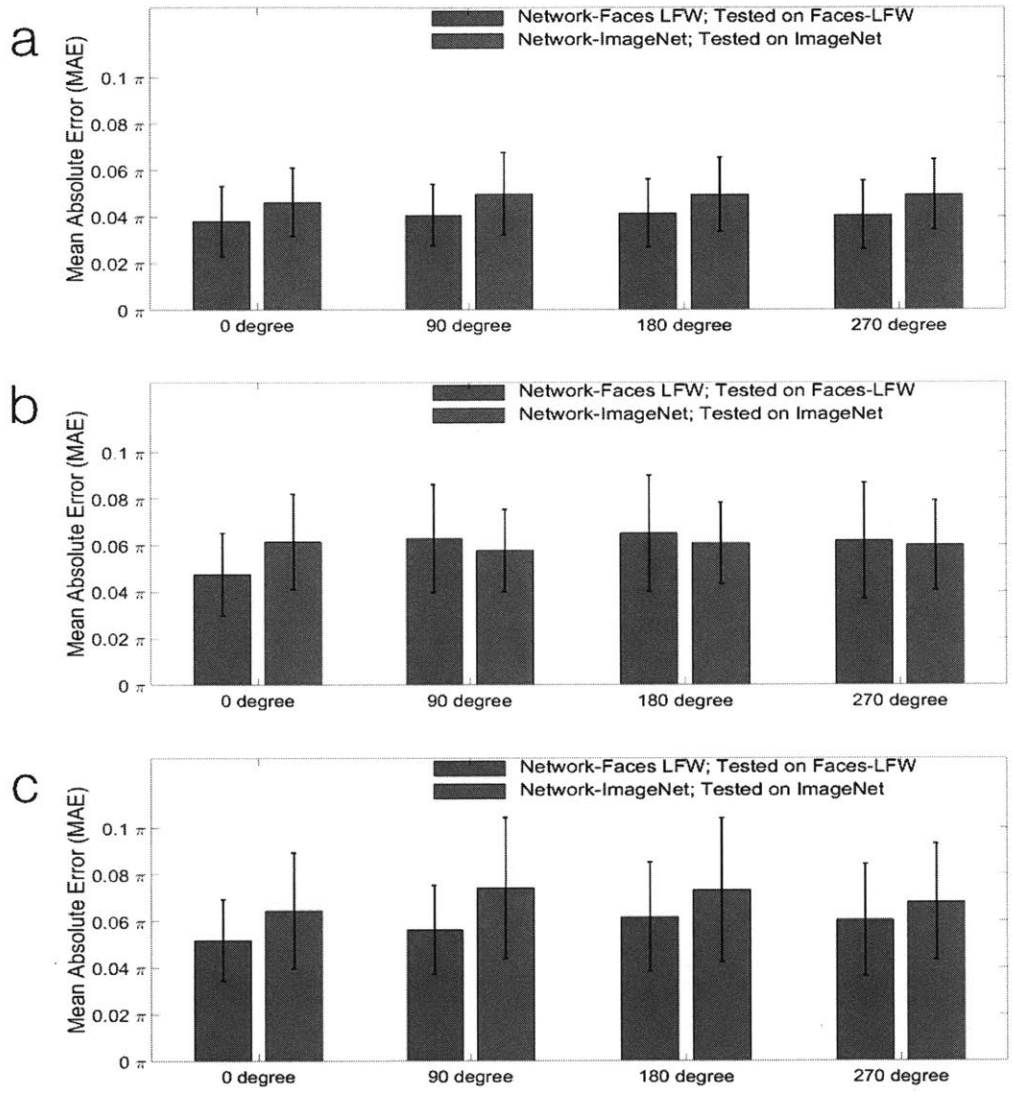

Figure 5-11: Quantitative analysis of the sensitivity of the trained deep convolutional neural network to rotation of images on the SLM. The baseline distance on which the network was trained is (a) z_1, (b) z_2 and (c) z_3, respectively.

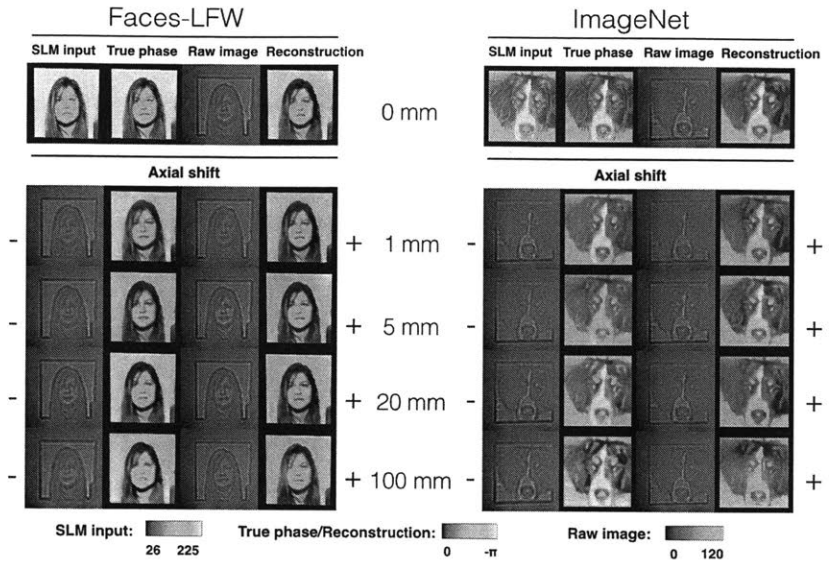

Figure 5-12: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to the object-to-sensor distance. The baseline distance on which the network was trained is z_1.

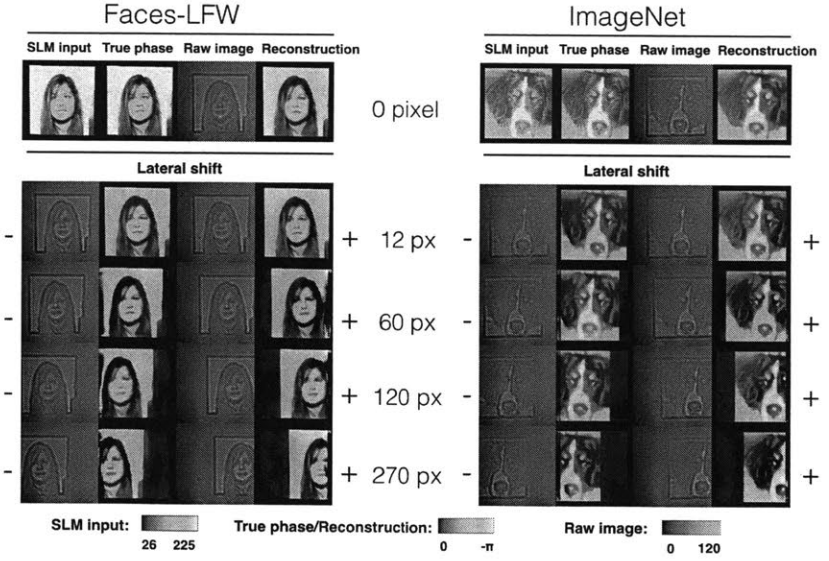

Figure 5-13: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to lateral shifts of images on the SLM. The baseline distance on which the network was trained is z_1.

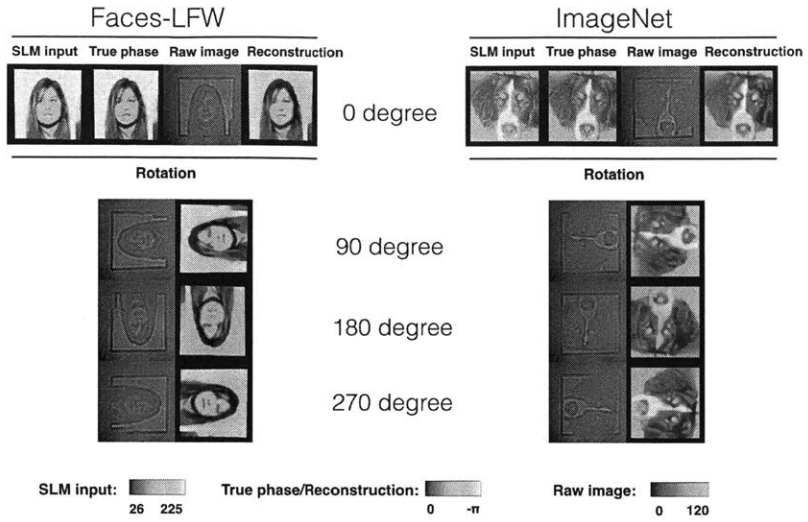

Figure 5-14: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to rotation of images in steps of 90. The baseline distance on which the network was trained is z_1.

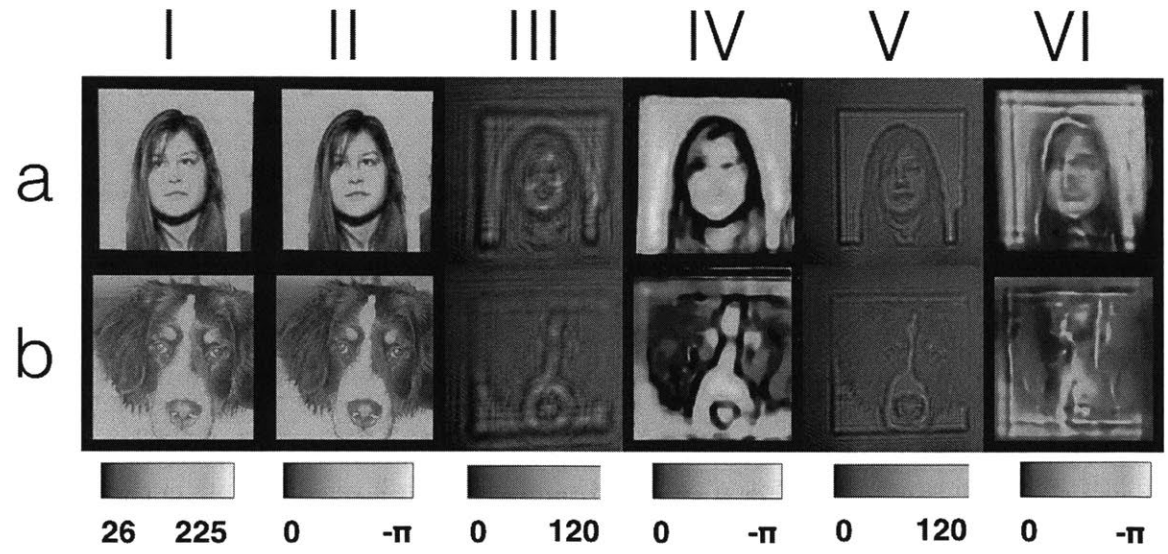

Figure 5-15: Failure cases on networks trained on Faces-LFW (row a) and ImageNet (row b) datasets. (i) Ground truth input, (ii) calibrated phase input to SLM, (iii) raw image on camera (iv) reconstruction by DNN trained on images at distance z_1 between SLM and camera and tested on images at distance 107.5 cm, (v) raw image on camera and (vi) reconstruction by network trained on images at distance z_3 between SLM and camera and tested on images at distance 27.5 cm.

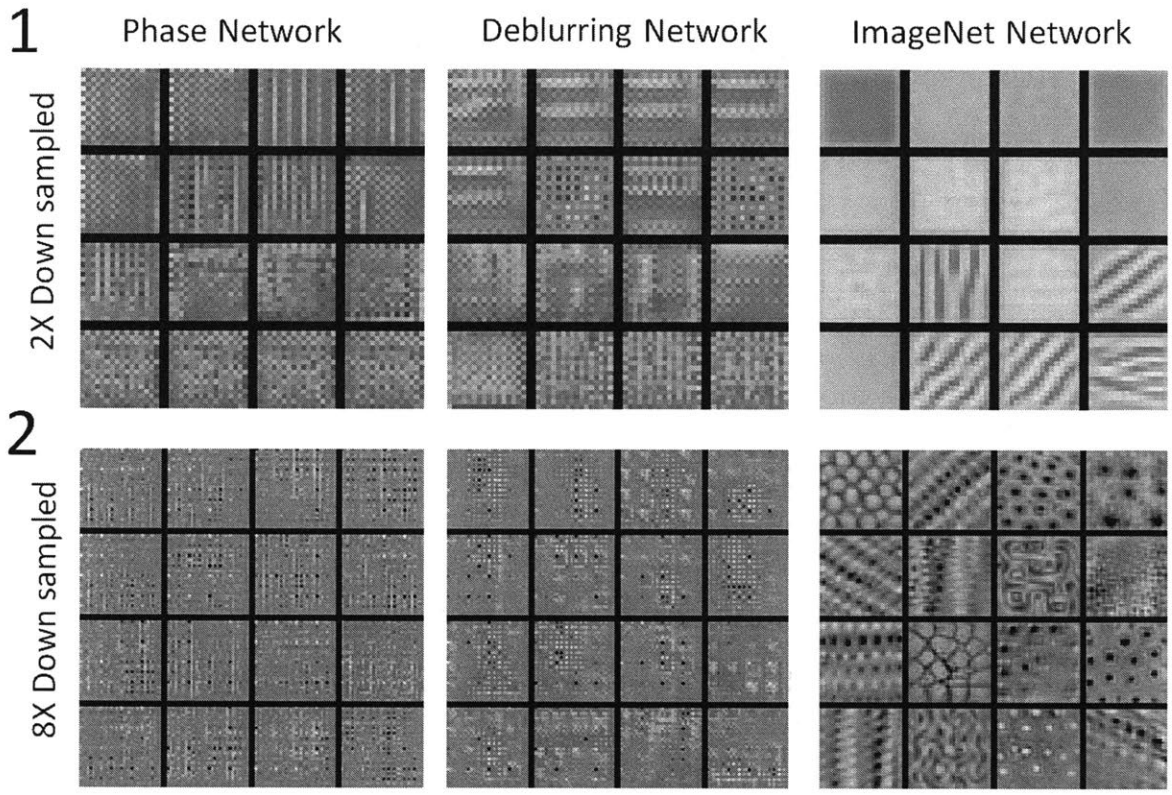

Figure 5-16: (1) 16×16 inputs that maximally activate the last set of 16 convolutional filters in layer 1 of our phase retrieval network trained on ImageNet at distance of z_1, a deblurring network, and an ImageNet classification network. The deblurring network was trained on images undergoing motion blur in a random angle within the range [0,180] degrees and a random blur length in the range [10,100] pixels. The image is downsampled by a factor of 2 in this layer. (2) 32×32 inputs that maximally activate the last set of 16 randomly chosen convolutional filters in layer 3 of: our network, the same deblurring network, and the ImageNet classification network. The raw image is downsampled by a factor of 8 in this layer.

Chapter 6

Computational Imaging for Mitosis Detection

The first section of this chapter reviews the theory of ptychography and its Fourier dual, Fourier ptychography. An experimental demonstration of Fourier ptychography for stained onion root tip imaging is shown, and a simulation demonstrating the potential of ptychography for imaging through turbid media is introduced and discussed.

The second section of this chapter documents a system we built for automated mitosis detection in a desktop brightfield microscope. The setup was first done in a brightfield arrangement and LED array illumination was later added to perform Fourier ptychographic imaging.

The third section details an experiment testing the hypothesis that deep learning mitosis detection from a focal stack of images (instead of a single in-focus image) can provide increased detection accuracy.

The fourth and final section presents the author's thoughts and ideas for future work in this domain.

6.1 Ptychography and Fourier Ptychography

6.1.1 Ptychography

Ptychography is a an imaging technique where an illuminating probe function (e.g., a physical aperture) is scanned over a sample and then propagated to an output plane where a time-averaged intensity is recorded [62, 63]. The term "ptychographiy" derives from the greek root $\pi\tau\upsilon\chi\hat{\eta}$ for "fold" and relates to the overlapping nature of the shifting probe function used in image acquisition (images are overlapped or "folded" over each other).

An example ptychographic imaging setup is shown in figure 6-1.

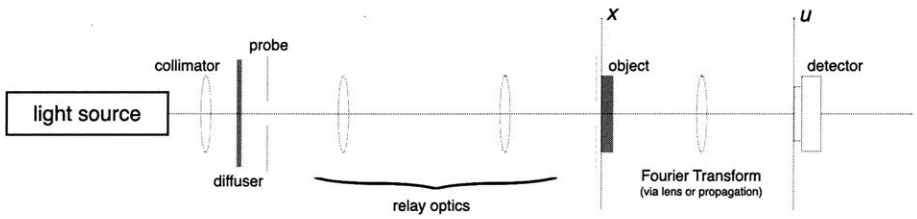

Figure 6-1: In ptychography, an illuminating probe function is scanned over a sample and propagated to an output plane.

Ptychographic images acquired from this setup can be written as:

$$I_{\text{out}}(u, s_0) = \left| \int p(x - s_0) \cdot o(x) e^{-j2\pi x u} dx \right|$$

where p is the probe function, o is the object function and s_0 represents a lateral shift of the probe function as it is scanned across the object.

Several algorithms exist for phase retrieval from a ptychographic dataset. One such algorithm, Wigner Distribution Deconvolution, is discussed in Chapter 3 of this thesis. Another, more popular, algorithm for ptychographic phase retrieval is the "extended Ptychographic Iterative Engine" (ePIE) [133, 134]. The ePIE uses alternating projections to converge to a solution consistent with all acquired images. A random initial phase is assigned to the initial object estimate and the ePIE proceeds as follows:

1. Current estimates for object $o(x)$ and probe $p(x)$ are forward-propagated, e.g., as $\mathcal{F}\{o(x) \cdot p(x - s_i)\}$), to estimate an output field $G_{\text{out}}(u)$

2. The amplitude of this estimated $G_{\text{out}}(u)$ is replaced with the experimentally observed amplitude

3. This new $G_{\text{out}}(u)$ is Inverse Fourier Transformed to obtain an estimate for $o(x) \cdot p(x)$

4. Gradient descent is applied twice to simultaneously update the object and probe estimate.

5. Steps 1-4 are repeated until convergence

An illustration of the ePIE in action is shown in Figure 6-2

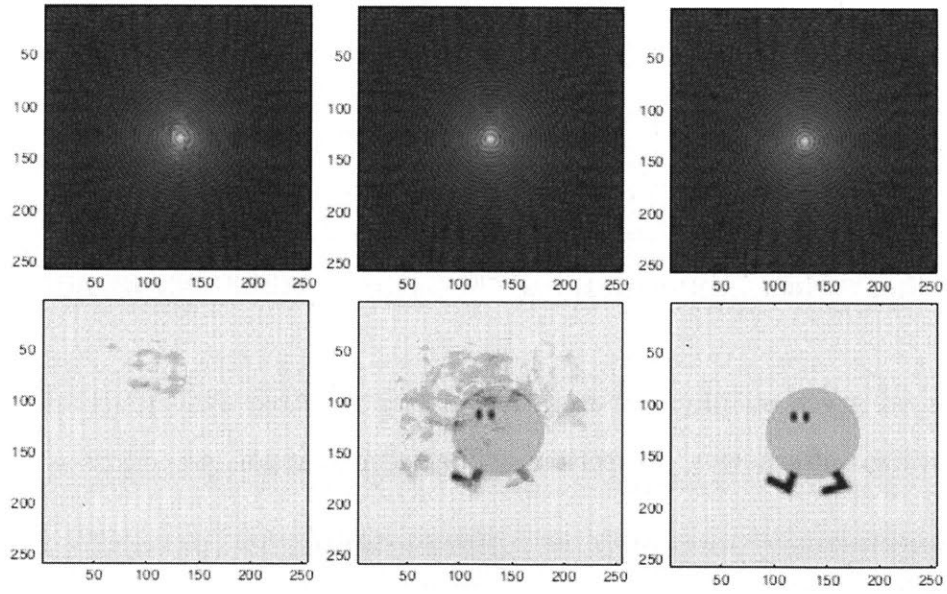

Figure 6-2: Phase retrieval using the extended Ptychographic Iterative Engine. By enforcing consistency in overlapping regions' measurements, both the probe function and object function can be simultaneously retrieved using the ePIE. (left) object estimate at early stage of ePIE algorithm, (middle) object estimate at later stage of ePIE algorithm (right) final object estimate from ePIE algorithm.

6.1.2 Fourier Ptychography

Fourier ptychography is the Fourier dual of ptychography; instead of shifting the object (or probe) in real-space and taking measurements in frequency space, shifts are performed in frequency space (by changing angle of illumination) and intensity measurements are recorded in the real-space domain [70]. An example Fourier ptychographic imaging setup is shown in figure 6-5.

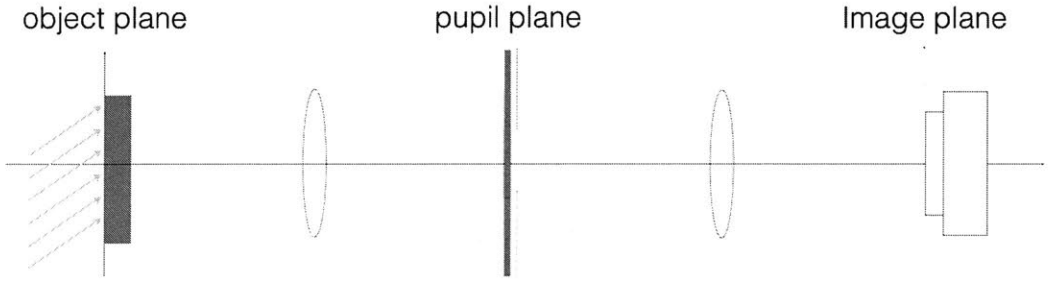

Figure 6-3: Fourier Ptychography Imaging Setup.

Mathematically, the images acquired from a Fourier Ptychographic imaging setup can be expressed as [70]:

$$I_{\text{out}}(x, s_u) = \left| \int P(u) \cdot O(u - s_u) e^{-j2\pi x u} dx \right|^2$$

where $O(u)$ is the frequency spectrum of the object, shifted by s_u, the spatial frequency of the incident illumination, and $P(u)$ represents the imaging system's pupil function.

Similar to ptychography, the iterative fourier ptychographic phase retrieval algorithm utilizes alternating projections. The reconstruction procedure is as follows [135]:

1. Plug in current estimates for object's frequency spectrum $O(u)$ & system pupil function $P(u)$ through the propagating function to obtain an estimate for the output field $g_{\text{out}}(x)$

2. Replace the amplitude of this estimated $g_{\text{out}}(x)$ amplitude w/ experimentally observed amplitude

3. Inverse Fourier Transform g_{out} to obtain estimate for $O(u) \cdot P(u)$ ("FT. of object .* ATF")

4. Use gradient descent twice to simultaneously update object frequency spectrum and ATF/pupil function estimates.

5. Repeat above steps until convergence

An illustration of the iterative reconstruction from Fourier ptychographic images is shown in Figures 6-5 and 6-4.

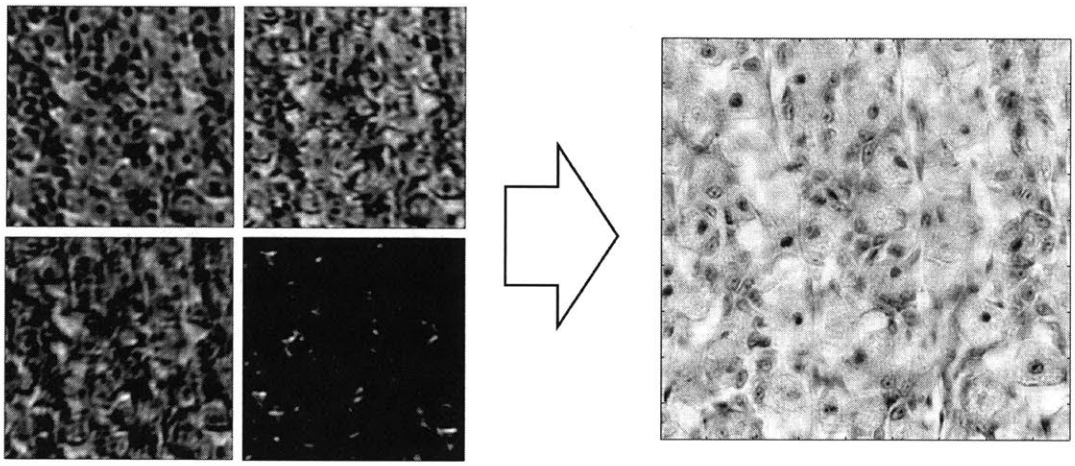

Figure 6-4: In Fourier Ptychographic Microscopy, many low resolution images (captured using different angles of illumination) are synthesized into a single higher resolution image

6.1.3 Experimental Fourier Ptychography

An experimental Fourier Ptychographic imaging system was used to image onion root tip slides. The system had magnification $M = 10.27$ and utilized an LED array with 4mm spacing placed 79mm above the sample to sequentially illuminate the sample with $\sim$ 630nm light. Sixty-nine low-resolution images were captured and used to reconstruct a higher resolution image using the extended ptychographic iterative engine (ePIE) algorithm. Calculations were performed on a Mac Pro, and for a set of 1024x1024 images, the ePIE algorithm converged within 30 seconds. Reconstruction results are shown in Figure 6-6.

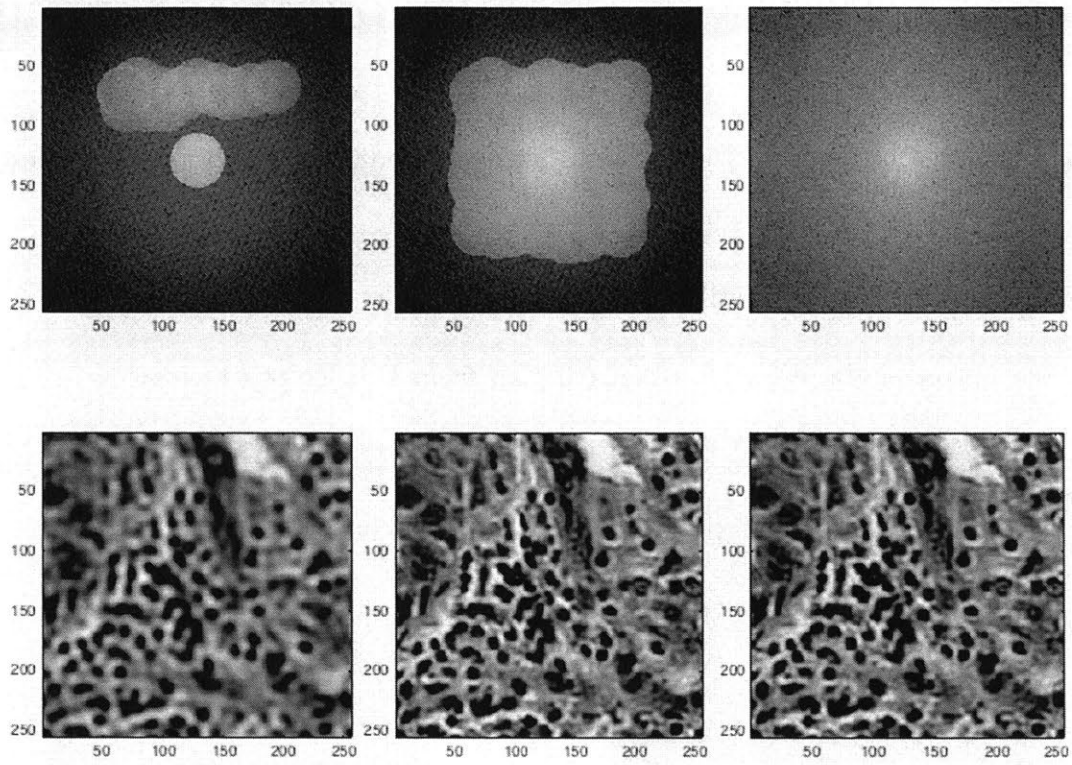

Figure 6-5: Fourier Ptychographic Image Reconstruction. (left) object estimate at early stage of algorithm progression, (middle) object estimate at later stage of algorithm, (right) "ground truth" image

Notably, although significant resolution improvement is observed between the on-axis low-resolution image and computationally reconstructed high-resolution image, neither image looks that great compared to the brightfield microscopy images obtained in Section 2 of this chapter. This is in large part due to the resolution of incoherent imaging systems being intrinsically double that of an identical coherent imaging system and in part due to coherence artifacts such as interference rings.

6.1.4 Fourier ptychographic imaging through turbid media

We simulated the experimental setup shown in Figure 6-7 to assess the ability of an ePIE-like iterative algorithm for imaging through turbid media. In our simulation, a high-scattering plane (P2) is placed some distance behind an object plane of interest (P1). Using a multi-slice iterative reconstruction algorithm similar to that demonstrated in [136] and [137], we were able to recover both the high-scatter plane and the

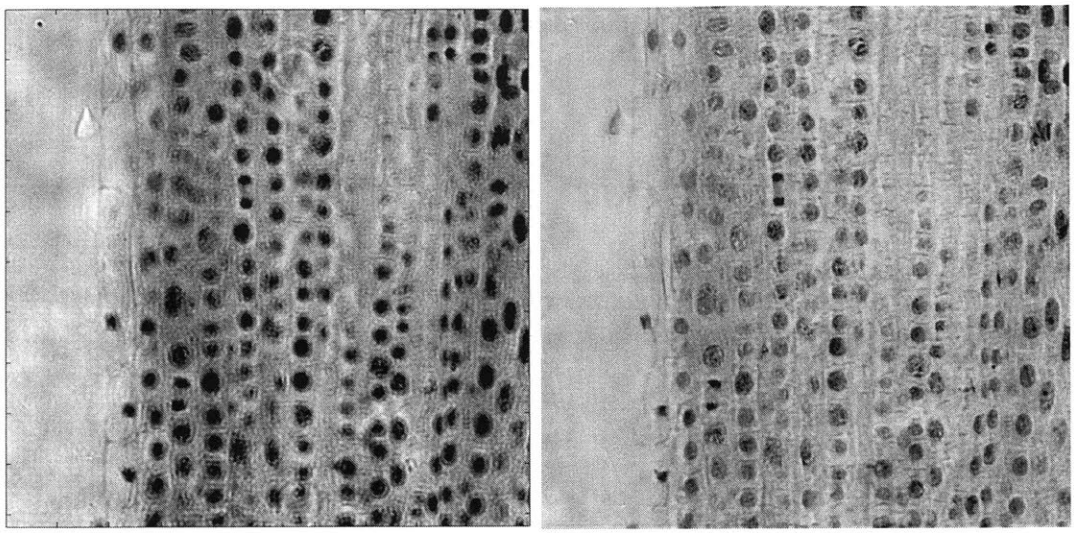

Figure 6-6: Fourier Ptychographic Image Reconstruction of onion root tip. (left) raw low-resolution on-axis image fed into ePIE algorithm, (right) high-resolution ePIE reconstruction

input plane. Results are shown in Figure 6-8. We then swapped P1 and P2 and found that the algorithm could still reconstruct both the object and scattering planes. Notably, this algorithm took significantly longer to converge vs standard ptychography for a single slice (hours vs minutes), suggesting that it may be much more sensitive to experimental error or misalignment.

6.2 MITOscope: Automated mitosis detection in a desktop microscope

We built and tested a system for automated mitosis detection in a desktop microscope. The platform used for this experiment was a low-cost (<$300) OMAX microscope and camera available from Amazon.

Since the author was unable to procure labeled pathology slides of cells undergoing mitoses, he used slides of onion root tip mitoses and self-labeled the dataset. Onion cells in any stage of mitosis (prophase, metaphase, anaphase, or telophase) were labeled as "positive" and remaining (unlabeled) parts of the image were considered

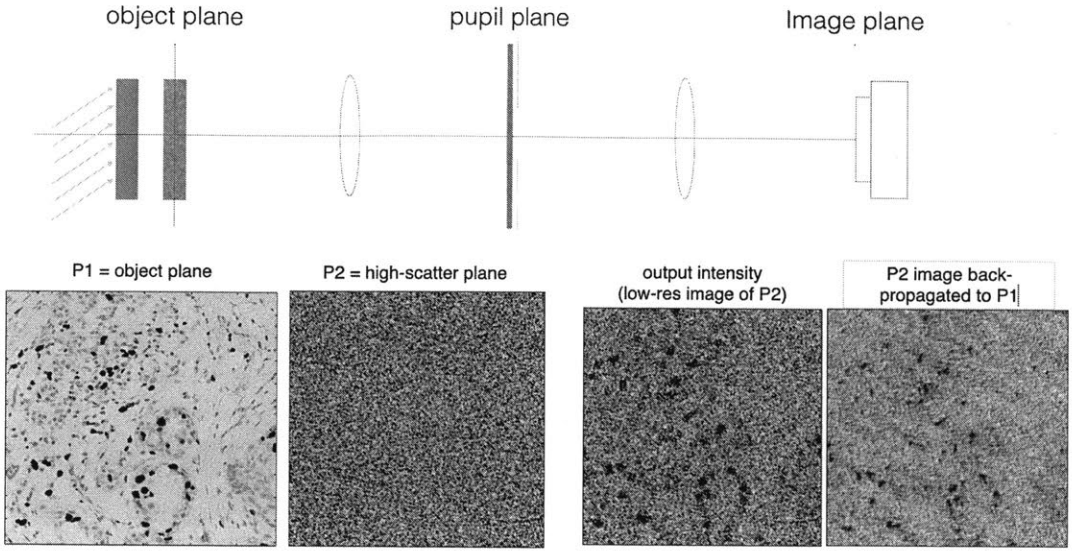

Figure 6-7: Fourier ptychographic imaging through turbid media.

"negative."

In practice, the resolution gain from doing Fourier Ptychography (Fig 6-6) did not exceed the 2x gain obtained via incoherent imaging. Additionally, due to a lack of coherence artifacts, images obtained via brightfield imaging were of higher quality than the images obtained via Fourier Ptychography, and for the remainder of this section we use images acquired from a brightfield setup at low (10x) magnification. The 10x objective used had an NA of 0.45, corresponding to a resolution of about 0.6 microns. A pathologist cannot count mitoses at this magnification, but optically, the information is there (provided that the camera sensor has an appropriate space-bandwidth product and pixel size). Thus, we used this setup to demonstrate the ability of our end-to-end imaging platform+mitosis detection algorithm for counting mitoses at low magnification across a wide field-of-view.

A DNN was trained as per the methodology in the following section of this chapter ("Deep learning mitosis detection from focal stack images"), and an example output of the trained classifier for a test/holdout image is shown in Figure 6-11. The trained DNN (at the operating point of the ROC) had a sensitivity ("true positive rate") of 0.913, precision ("positive predictive value") of 0.944, and an F1 score of 0.928 (220 TP, 21 FP, 13 FN). Notably, the classifier actually identified 17 mitoses that had been

missed during the labeling process. These were not counted as false positives for the above analyses, and it is likely that if the network were retrained with these examples correctly labeled, the network could achieve significantly improved results.

In our benchtop setup, images were acquired at 10x magnification in less than a second, and running inference over a single image took about 30 seconds per image. The inference step is highly parallelizable and could be sped up with more GPUs to process an entire image is less than a second. Based on the demonstration reported here (for onion cell mitoses), we envision a similar benchtop microscope system for pathologists where mitoses (or any other hallmarks of disease severity) can be located and counted "automagically" from low-magnification (high field of view) images at the push of a button. The benchtop nature of the setup allows pathologists to rapidly switch between high and low magnification to double-check the algorithm's results, and the added benefit of being able to "see the bigger picture" of mitoses (i.e., at wide field of view) may facilitate and improved understanding of a patient's disease.

6.3 Deep learning mitosis detection from focal stack images

In this section, we test the hypothesis that using a focal stack of images instead of a single in-focus image as input can significantly improve the accuracy of these deep learning algorithms. If our hypothesis is true, it would suggest that given a limited photon budget (image acquisition time), it would be better to spend time acquiring a focal stack vs longer exposure times, even at the cost of increased noise/reduced SNR.

6.3.1 Dataset/Methods

Finding a labeled focal-stack dataset proved difficult since most whole-slide imaging systems discard out-of-focus images and few freely available online databases contain focal stack information. The one dataset we were able to find (and access) was

from the MITOS 2012 competition for multispectral imaging evaluation [138]. The intended purpose of the multispectral focal stack dataset was to evaluate whether or not multispectral images could improve machine learning results, and most contestants simply discarded all out-of-focus images in their analyses.

Here, we revisit the MITOS 2012 multispectral image dataset, which contains fifty 40X magnified Fields of View (FOV) from five different slides (10 per slide). Each 40X FOV is $512\mu m$-by-$512\mu m$ with pixel spacing of roughly 200nm. Image focal stacks were captured with z-plane separation of 500nm. In total, the multispectral dataset contained 170 gray-scale images (10 spectral bands, with a 17 layer z-stack for each spectral band). For our experiments, we picked just three color bands close to "R","G," and "B" (as would be captured by standard digital camera) to simulate what an image under a standard (non-hyperspectral) imaging system would look like. These bands were: band 3 ("B"; 430nm-490nm), band 6 ("G"; 525-570nm), and band 0 ("R"; 665-710nm). Additionally, due to computational constraints, we selected just three images from the focal stack to use for DNN training/evaluation: an in-focus image, an image overfocused by $2.5\mu m$ and an image underfocused by $2.5\mu m$.

To choose these three focal stack images, we first scanned through all of the images in the focal stack and estimated an in-focus plane/image using a sharpness metric (similar to classical in-line digital holography [145]). Then, we selected this in-focus image/plane and two images/planes that were $\pm 2.5\mu m$ above or below that focal plane to use in our dataset(s). Since the numerial aperture of the whole slide scanner used for this dataset was close to 1.0 (assuming that optical resolution of the WSI system was correctly matched with its pixel resolution of 200nm), the selected out-of-focus images were well outside the depth of field of the imaging system.

Since we wanted to be "photon-budget" fair when comparing our focal stack images against single in-focus RGB images, we synthetically added noise with varying levels of variance to our images in order to simulate various levels of exposure (e.g., for 1/3 the exposure time, noise variance was increased by 3x). From the originally provided 8-bit RGB images, noise with variance of 1,3,9,and 27 was added. For subsequent analyses RGB images were compared against focal stack images with 3x the noise

variance (e.g., var=3 vs var=9)

The generated noisy images were then fed into a DNN as 3-channel (RGB) or 6-channel (RGB+dI/dZ) inputs for training. "Lateral flux" (dI/dZ) inputs were obtained by differencing and normalizing the over and underfocused RGB images.

6.3.2 Training

We used a two-pass "boosting-inspired" DNN approach proposed by Ciresan et al [139] to train our classifier. This approach has also been successfully utilized by others [140] for tumor segmentation.

First, an initial training set ("training set A") is generated using cropped images centered at all mitoses pixels and an equal number of randomly selected nonmitoses pixels.

Then, a simple DNN was trained using training set A and used to run inference over the original labeled image (to generate a "mitosis-likelihood" probability map). Because the negative examples from Training Set A were more likely to contain images of areas that were not cell nuclei at all, this first-pass DNN had a very high false-positive rate.

Next, a second training set ("training set B") was generated using cropped images centered at all mitoses pixels and an larger number of nonmitoses pixels selected from a weighted probability distribution proportional to the probability map generated by the DNN trained using training set A. In this manner, training set B was comprised of more "difficult non-mitoses" examples than Training Set A, and the DNN could focus its efforts on learning to distinguish between true mitoses and difficult non-mitoses instances (instead of trivial examples of non-mitoses instances).

After training set B was generated, a new DNN was trained using set B and used to run inference over the training, test, and holdout images. An example output of this algorithm is shown in Figure 6-17. Before testing our entire pipeline on the multispectral dataset, we ran images from the non-multispectral MITOS2012 dataset through our pipeline to make sure the network architecture used was suitable (LeNet-like) for mitoses detection from in-focus color images. Using the Aperio (non-multispectral)

images from the MITOS2012 dataset to train our DNN, we obtained an F-score of ~ 0.767, which is consistent/competitive with results from literature [139].

6.3.3 Results

We passed our multispectral datasets through the pipeline above and stopped computation just before running final inference across the entire images (due to time/resource constraints) for initial comparison. In order to validate our results, we ran both 4-fold cross-validation and Monte-Carlo cross-validation. Our results are shown in Fig. 6-18.

Unfortunately, our results suggest that multifocal imaging, at least in the implementation presented here, does not significantly improve mitosis classification accuracy. Although on average, the focal-stack-trained DNN outperformed the RGB-trained DNN, the result was not significant and there were even some spits where the RGB-trained DNN outperformed the focal-stack-trained DNN. It is likely that the spatially coherence of the illumination in the multispectral imaging system was too low for the out-of-plane images to contain any scattering/phase information from the sample, and a repeat of this experiment in a coherent or partially coherent illumination setup is as suggested subject of future study.

Another takeaway from our experiment was that the trained DNNs were very robust to noise; (no correlation between noise level and classification accuracy). This was not too surprising since DNNs are known to be robust to noise and jitter is often added as a means to augment training data.

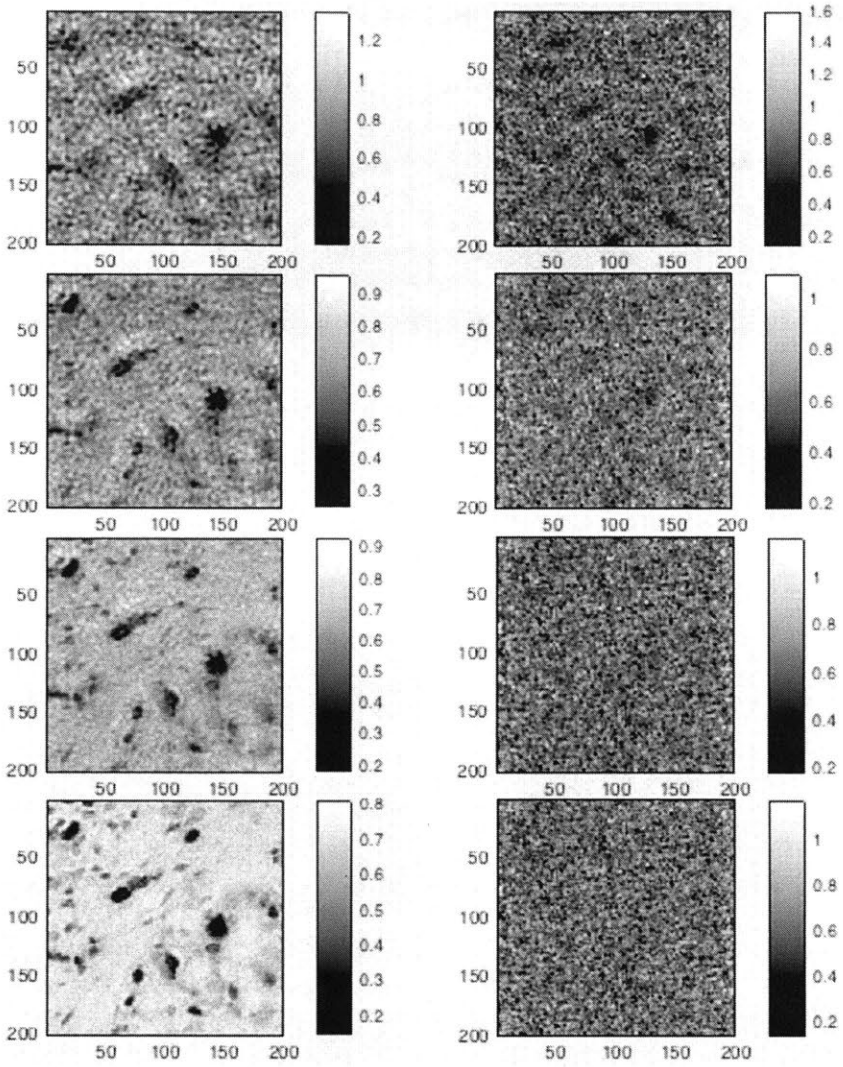

Figure 6-8: Fourier ptychographic imaging through turbid media. Top-to-bottom image pair progression shows the algorithm's current estimates for P1 (left) and P2 (right) over time.

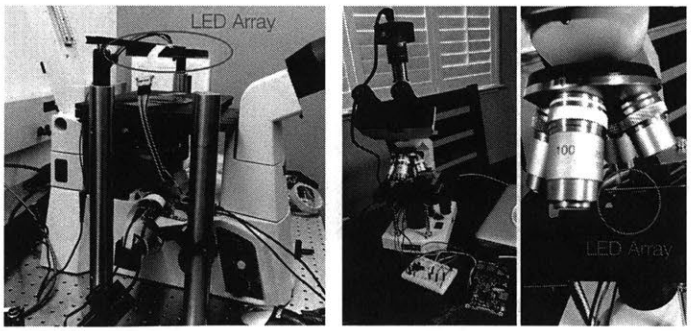

Figure 6-9: Example MIToscope setup. In the setup on the right, an LED array was inserted into the standard brightfield setup to allow Fourier ptychographic imaging.

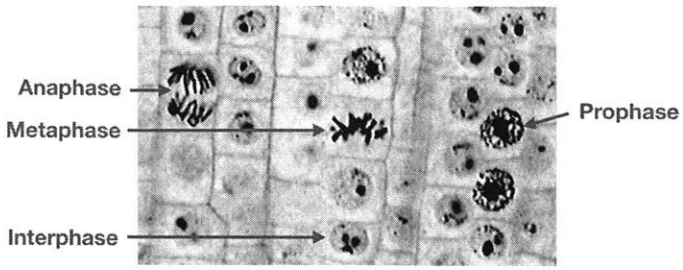

Figure 6-10: Onion cells in various states of cell cycle.

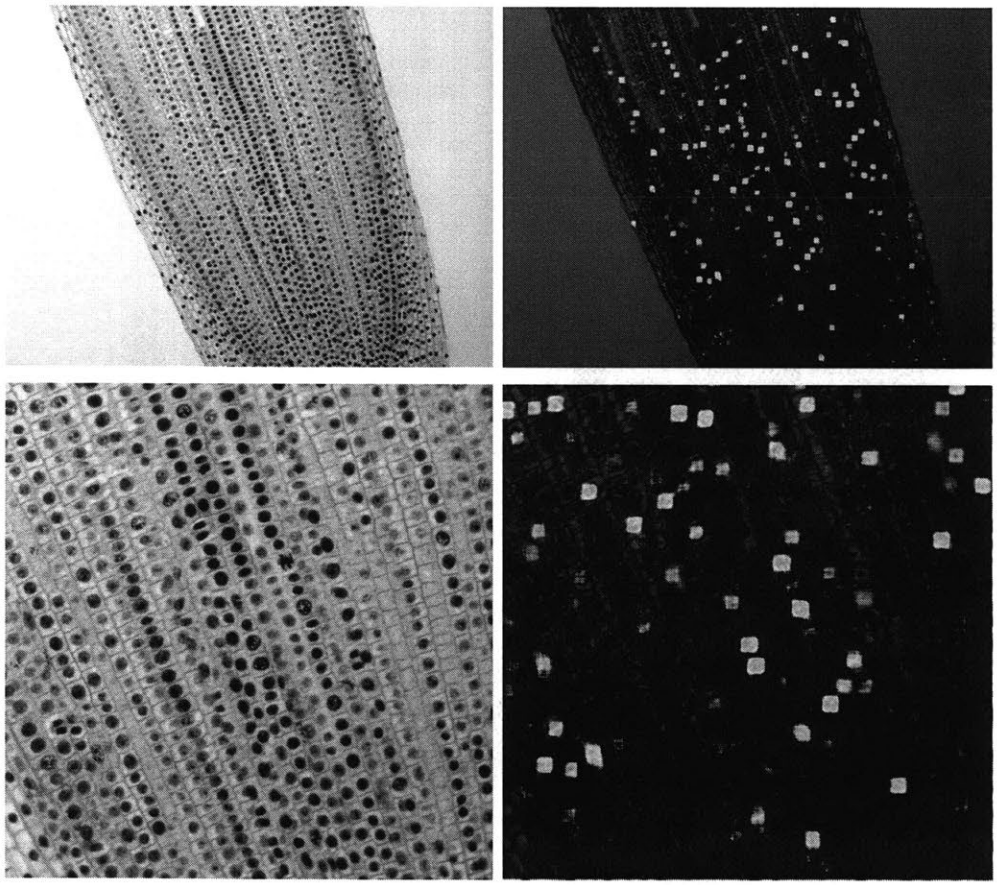

Figure 6-11: Automated onion cell mitosis detection. (top) full field of view/single image (bottom) zoomed-in field of view

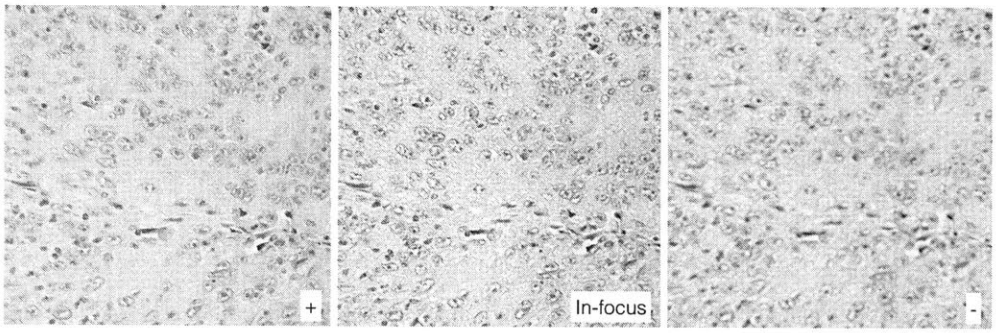

Figure 6-12: 2.5μ over-focused, 2.5μ underfocused, and in-focus images were used to train our DNN.

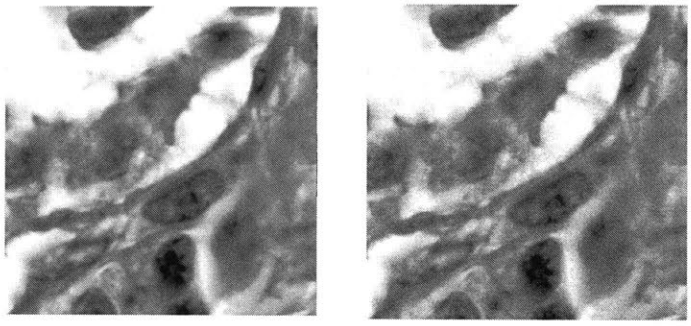

Figure 6-13: Noise added to images to simulate equal photon budget.

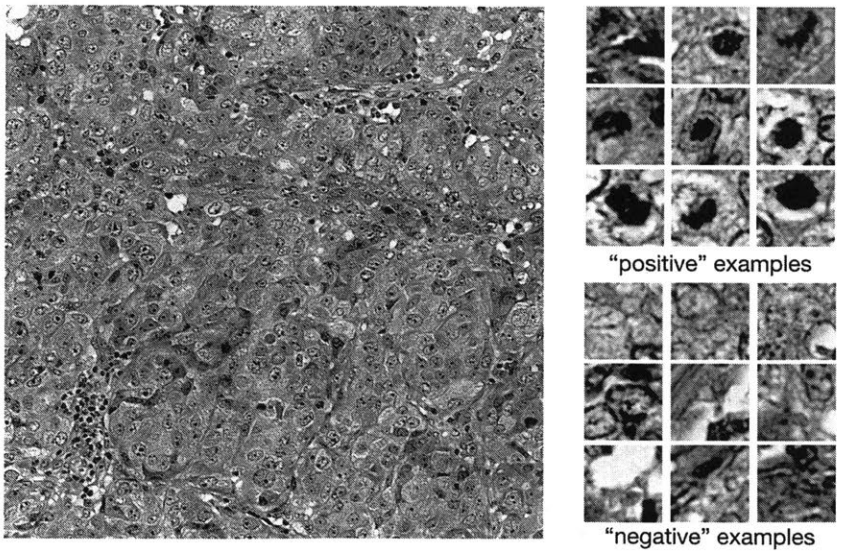

Figure 6-14: First training set drawn from labeled image. (top-right) cropped images centered at mitoses pixels (bottom-right) randomly selected non-mitoses pixels

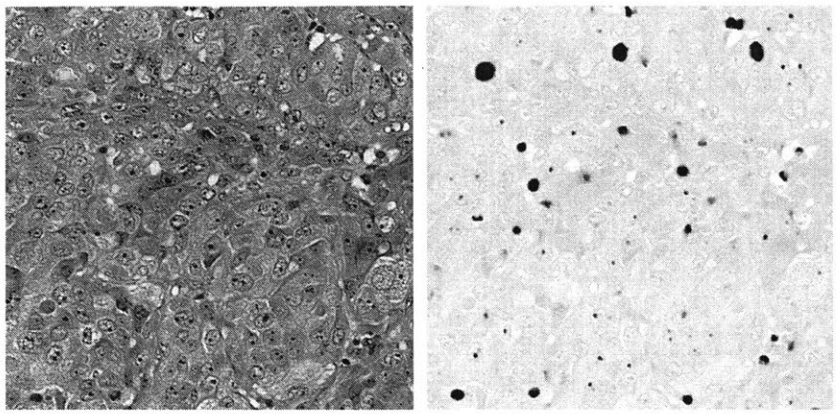

Figure 6-15: The probability map generated by a DNN trained using set A has a high false-positive rate. (left) ground truth image with mitoses labeled in yellow. (right) mitosis probability map (darker = higher probability)

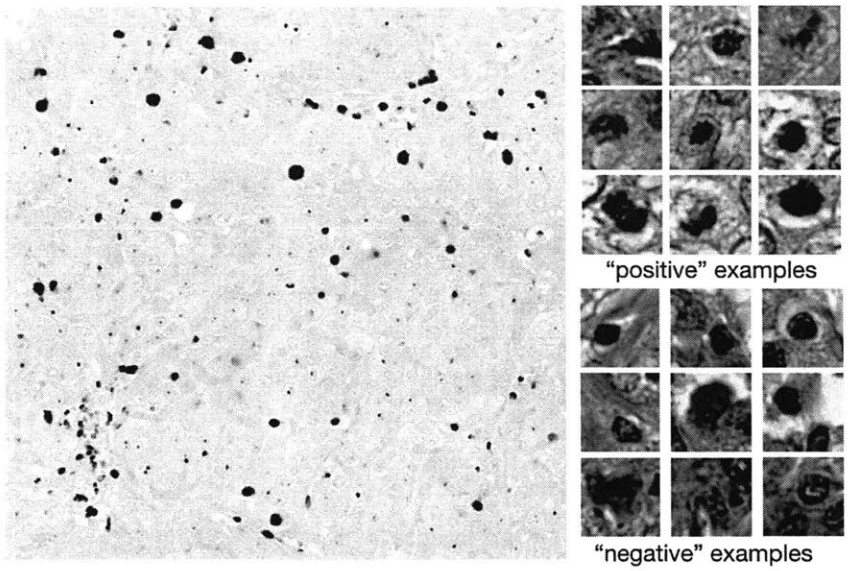

Figure 6-16: Training set B was generated using cropped images centered at mitoses pixels (top right) and non-mitoses (bottom right) pixels selected from a weighted probability distribution proportional to the probability map generated by the DNN trained using training set A (left)

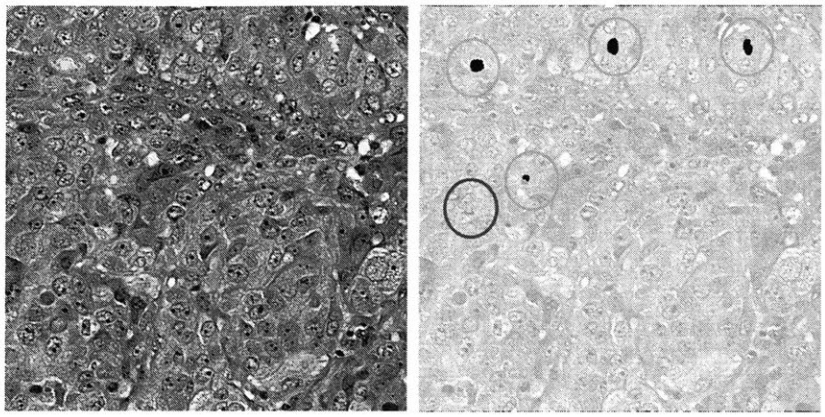

Figure 6-17: (left) ground truth image with mitoses labeled in yellow. (right) Example DNN output when trained using training set B. Green circles denote correctly identified mitoses; the red circle indicates a missed mitoses (false negative)

Test-set results: Accuracy

Noise Variance	1	3	9	27
RGB_stack	0.916	**0.912**	0.915	**0.913**
RGB	**0.898**	0.907	**0.912**	0.899

Monte Carlo cross-validation (16 randomized training/test splits)
[binary cross-entropy loss]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
RGB_stack var=9	0.229	0.280	0.270	0.256	0.262	0.231	0.228	0.259	0.231	0.218	0.279	0.241	0.278	0.239	0.255	0.226		
RGB var3	0.260	0.274	0.277	0.240	0.263	0.252	0.264	0.280	0.249	0.249	0.254	0.257	0.252	0.263	0.257	0.254	**Avg Diff**	**StDev**
	0.031	**-0.006**	**0.007**	**-0.016**	**0.001**	**0.021**	**0.036**	**0.021**	**0.018**	**0.031**	**-0.025**	**0.016**	**-0.026**	**0.024**	**0.002**	**0.028**	**0.010**	**0.009**

Figure 6-18: Mitosis detection results.

Chapter 7

Concluding Thoughts

This thesis is part of a larger body of efforts to apply computational imaging to pathology. This new way of thinking about imaging for pathology will enable pathologists to image faster, cheaper, and more robustly. Images reconstructed via computational imaging are not limited by the physics of traditional imaging systems, and computational imaging platforms can image beyond the diffraction limit at a much higher space-bandwidth product than brightfield microscopy. Moreover, these images contain a much richer set of information (e.g., quantitative phase or phase-space) which can be used to synthesize extended depth-of-field images in thick samples or provide extra information to machine learning algorithms to help improve model accuracy.

Machine learning is eating the world, and pathology is no exception. State-of-the-art deep learning algorithms can already outperform pathologists at several tasks including tumor segmentation and mitosis detection. In the future, AI will play an increasingly greater role in personalized medicine, predicting treatment response and guiding therapeutic decisions.

What role does computational imaging play in this future? As the field trends towards AI-assisted diagnoses via digitized whole-slide images, computational imaging can help fill the need for high throughput imaging systems which can image whole slides quickly, robustly, and cheaply. Additionally, it is helpful to keep the end-goal of such imaging systems in mind when designing computational imaging platforms for high-throughput whole-slide imaging (WSI). In particular, the WSI system need

not be optimized purely for "quality of image reconstruction," but instead can also be optimized for clinical result or utility.

We are excited by the immense potential of of computational imaging to transform the field of pathology. Ongoing efforts stemming from the work done in this thesis include: a deep learning system for imaging through turbid media (e.g., skin or blood) [141], applications of the system developed in Chapter 4 for imaging biological specimens [142, 144], and a high-speed whole-slide imaging platform.

Beyond pathology, another field touched upon in (but not an explicit focus of) this thesis is the application of machine learning to computational imaging itself. For decades, advances in computational imaging have stemmed from clever human-developed algorithms taking advantage of underlying physical models of light propagation or *a priori* assumptions about the object(s) being imaged. Machine learning is eating the world, and computational imaging is no exception. We demonstrate in Chapter 4 of this thesis that a deep learning algorithm can outperform many of these hand-crafted models or algorithms. Deep learning has the potential to transform the entire field of computational imaging, and teaching machine learning algorithms how to "learn physics" is a nascent field of study that we predict will have enormous implications for the future of imaging.

Appendix A

Supplemental Material

This appendix provides supplementary information to "Lensless computational imaging through deep learning"

A.1 Calibration

We calibrated the HoloEye SLM using the Michelson interferometer experimental arrangement shown in Figure A-1 with M2 as the reference mirror. Lenses L1, L2 form a telescope that images the SLM onto the CMOS camera. We carried out two calibration experiments, one with all pixels in the SLM driven by the same value (uniform) and one with the left-hand half pixels driven by a constant value zero and the right-hand half driven by a different value (half-half.)

Let ϕ_0 denote the relative phase between the light reflected by the uniform SLM when its pixels are all driven with gray scale value of $V = 0$, relative to M2; and let $\phi(V)$ denote the phase when the SLM pixels are uniformly driven from $V = 1$ to 255. Let $\sqrt{I_0}$ denote the amplitude reflected by M2, and let $a(V)\sqrt{I_0}$ denote the residual amplitude modulation by the SLM, also as function of pixel values V. In the uniform illumination case, the intensity recorded by the CMOS should be of the form

$$\frac{I(V)}{I_0} = 1 + a^2(V) + 2a(V)\cos\phi(V) \tag{A.1}$$

We used the half-half version of the experiment, blocking the mirror M2 so that the beam on the CMOS is coming from the SLM only, to determine $a(V)$, and then determine $\phi(V)$. We also arbitrarily assigned $\phi(0) = 0$ radians. The two curves are shown, respectively, in Figures A-2 and A-3. The maximum intensity modulation ratio measured was ~ 1.3 for our configuration of polarizer and analyzer. In contrast, per manual specification, the SLM's maximum intensity modulation ratio is ~ 12000 for other polarizer/analyzer configurations. Since our intensity modulation is small, we neglected it while designing our neural network architecture.

Figure A-4 shows the variation of phase modulation with 8-bit gray scale values and three piecewise linear segments fitted to the phase modulation curve to obtain phase values varying linearly with gray scale values. The mean square error (MSE) of the fit is 0.18. In Figure A-5 we fit a single linear segment to the phase modulation curve with MSE of 0.19. All experiments in the main manuscript were done by training neural network with phase values obtained by fitting 3 segments to the curve. We report the results of phase retrieval by training networks on phase obtained by fitting a single linear segment in Section 3 of this supplement.

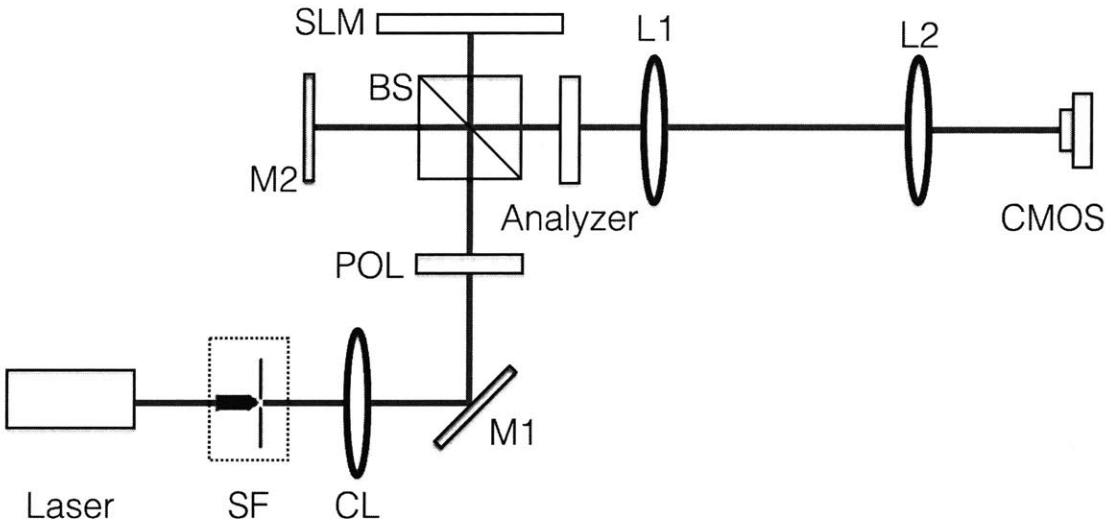

Figure A-1: The optical setup for calibrating the phase and intensity modulation of SLM. SF: spatial filter; CL: collimating lens; M1, M2: mirror; L1,L2: lens; POL: linear polarizer; BS: beam splitter; SLM: spatial light modulator.

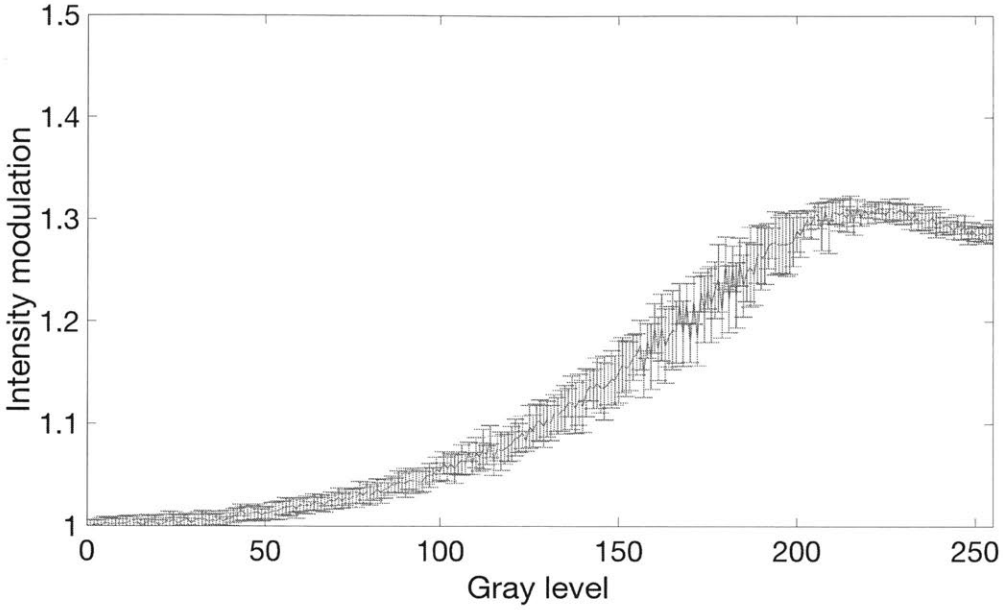

Figure A-2: Experimentally calibrated intensity modulation curve with error bounds in the grayscale range of [0,255] for the SLM.

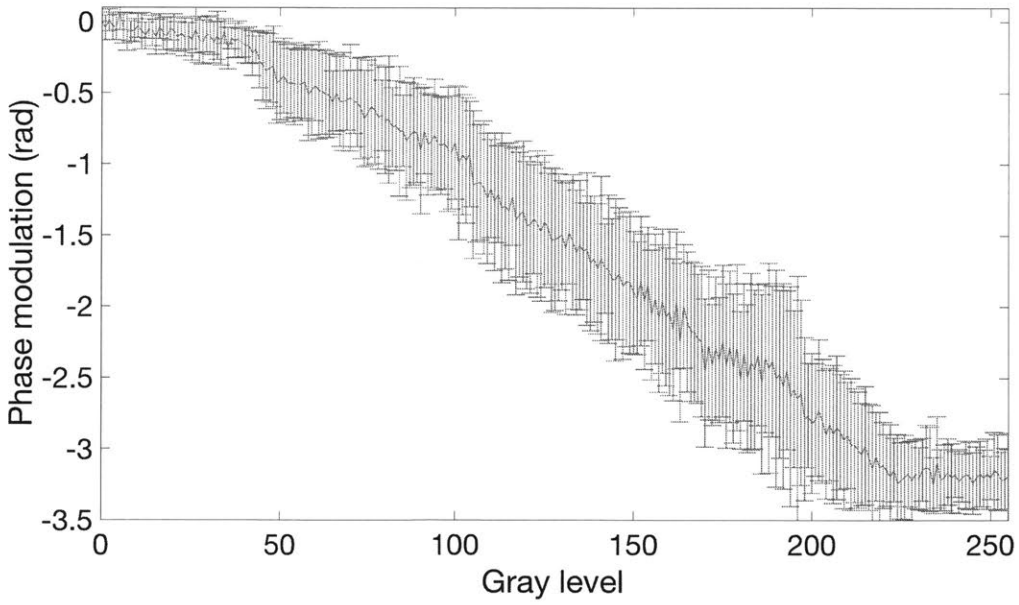

Figure A-3: Experimentally calibrated phase modulation curve with error bounds in the grayscale range of [0,255] for the SLM.

A.2 DNN Training

As described in the main manuscript, we used residual blocks to construct our DNN architecture. Here we describe additional details of our architecture. Figure A-6 shows

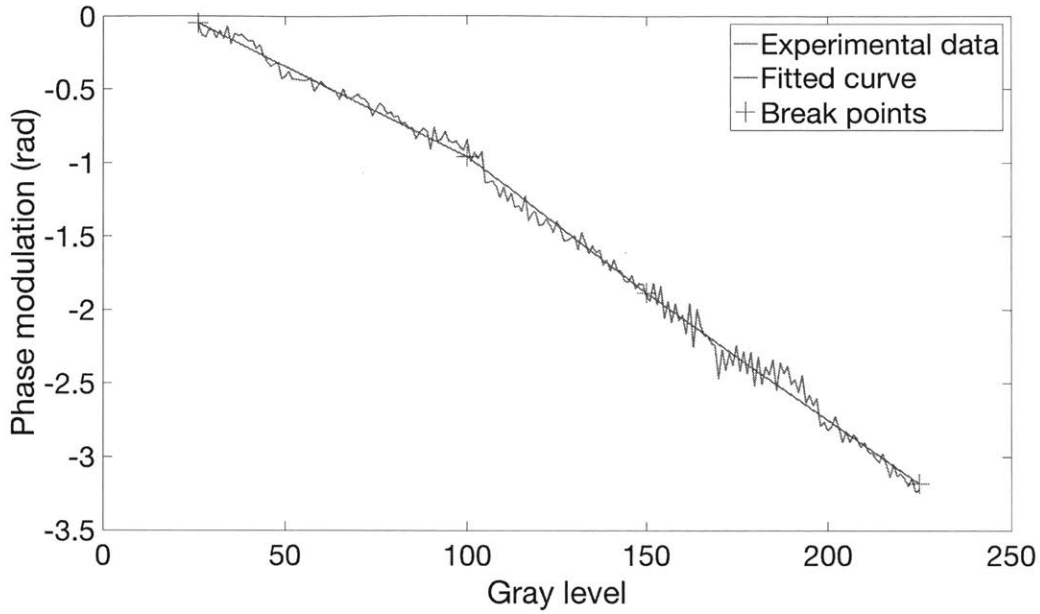

Figure A-4: Phase modulation curve along with three linear segments fitted to the curve.

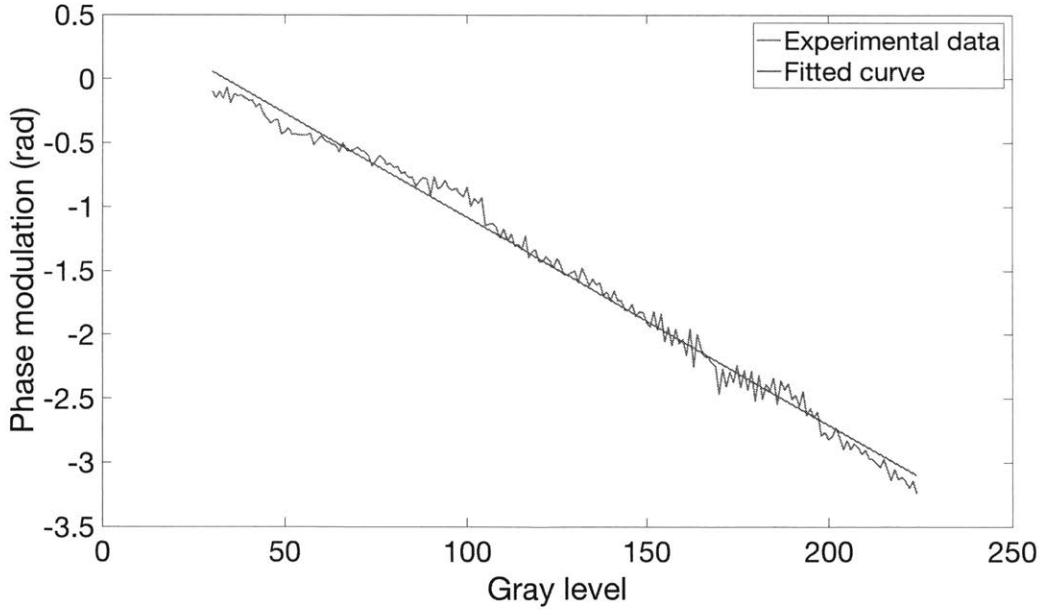

Figure A-5: Phase modulation curve along one linear segment fitted to the curve.

the different kind of residual layers (bottom row) composed of residual blocks (top row). All residual blocks are composed of 2 sets of batch normalization, nonlinearity (ReLU) and a convolutional layer stacked one above the other. All convolutional

filters are of size 3×3 and are either simple convolutions, convolutions with stride, convolution transpose or dilated convolutional filters depending on the type of the residual layer. The strides for convolution and convolution transpose filters, which downsample or upsample the input respectively, are set to $(2, 2)$ along the two spatial dimensions. Skip connections are either (a) direct connections that sum the input and output for blocks that do not change the size of the input, (b) 1×1 convolution filters with stride $(2, 2)$ for residual downsampling blocks, or (c) 2×2 convolution transpose filters with upsampling rate $(2, 2)$ for residual up-sampling blocks. We used weight decay of 1E-4 in all convolutional filters initialized with random numbers from a Gaussian distribution, and a dilation rate of 2 in all dilated convolutional filters. All residual layers are composed of two residual blocks. In our experiments, we observed that residual layers composed of 3 residual blocks (instead of 2) resulted in slightly lower mean absolute error (MAE) but much longer training times.

We set a small dropout rate of 0.02 between all layers to prevent overfitting. We used the ADAM optimizer in Tensorflow to minimize the MAE over the 10000 training samples with batch size 3. The batch size was constrained by the memory available on the GPU. We start the training with a learning rate of 0.001 and drop it by a factor of 2 after every 5 epochs. Additionally, we clip the gradients at value 1 to stabilize the training. We trained the neural network for 20 epochs shuffling the training samples at every epoch.

A.3 Axial perturbation, lateral perturbation and rotation at other two training distances

Figures 9, 10 and 11 in the main manuscript qualitatively showed the results of the DNNs (Faces-LFW and ImageNet) trained at distance $z_1 = 37.5\text{cm} \pm 2\text{mm}$ to axial, lateral and rotational shifts, respectively. Here we show the qualitative results of the sensitivity of the DNNs (Faces-LFW and ImageNet) trained at distances $z_2 = 67.5\text{cm} \pm 2\text{mm}$ and $z_3 = 97.5\text{cm} \pm 2\text{mm}$ to axial, lateral and rotational shifts.

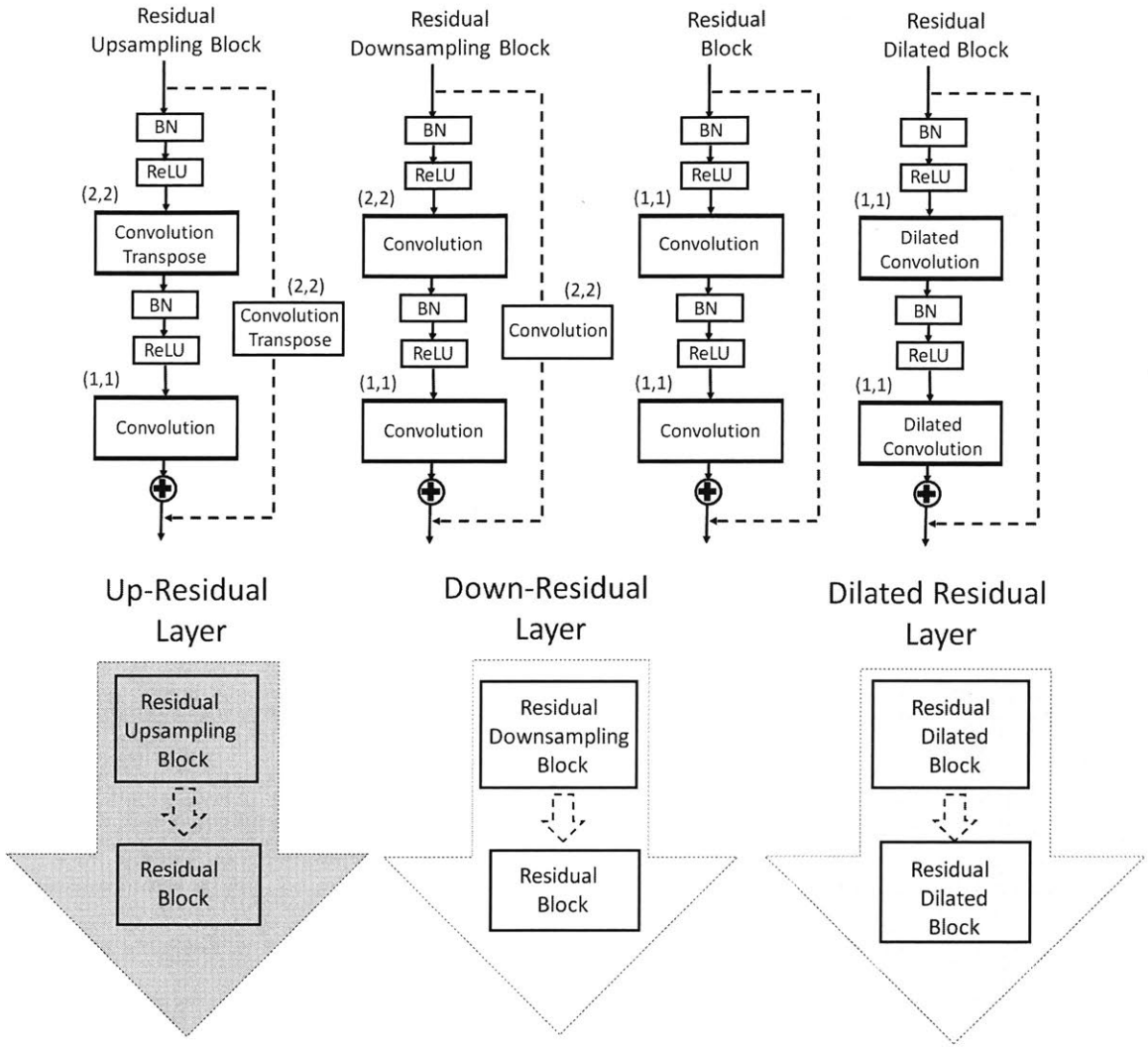

Figure A-6: Different types of residual layers used in our DNN architecture are shown in the bottom row which are composed of residual block structures described above. The strides for convolution filters in the residual blocks are shown above the filter.

Figures A-7, A-8 show the effect of changing the object-to-sensor distance relative to the distance the neural network was trained on, i.e., z_2 and z_3, respectively. 0 mm indicates the baseline training distance, and the sensor was displaced by 1mm, 5mm, 20 mm, and 100 mm both in the positive (further from the object) and negative (closer to the object) directions from the baseline distance. We see that the networks trained on Faces-LFW and ImageNet databases are both able to faithfully reconstruct the phase object even at large deviations from the training distance.

Figures A-9, A-10 show the effect of laterally shifting images on the SLM relative

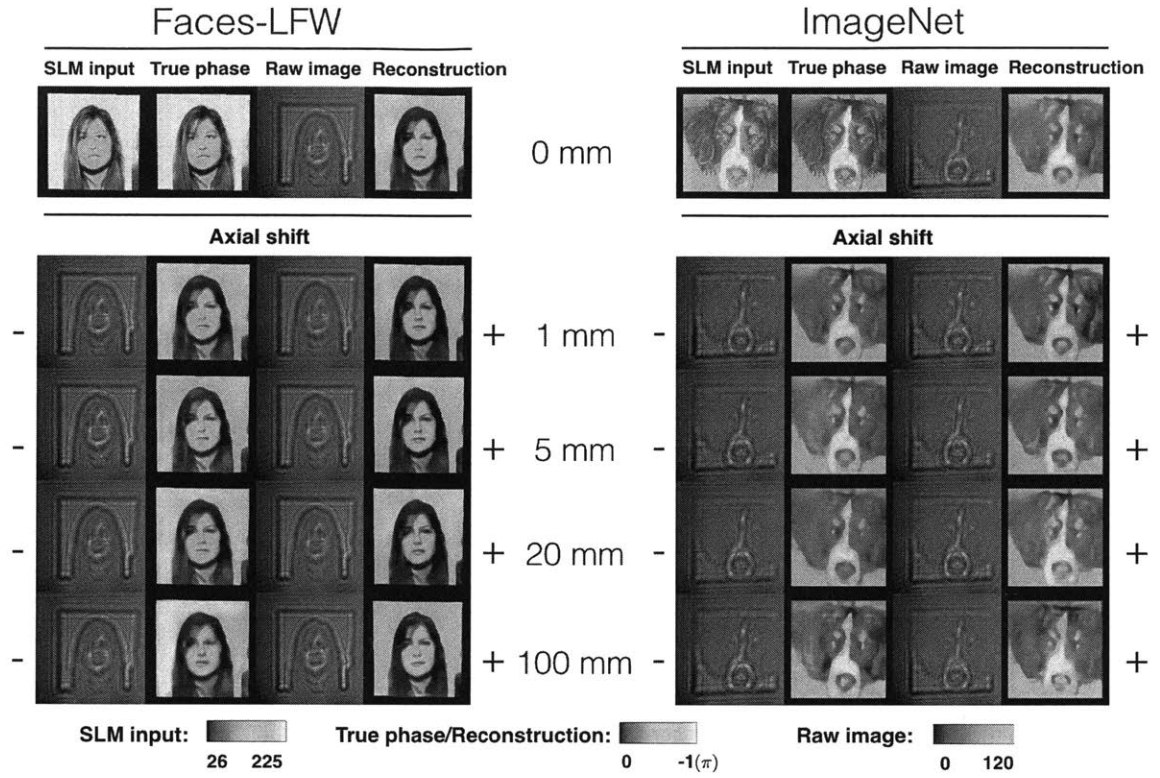

Figure A-7: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to the object-to-sensor distance. The baseline distance on which the network was trained is z_2.

to the baseline placement of the images on the SLM on which the neural network was trained at distances z_2 and z_3, respectively. 0 pixel indicates the baseline image placement used during training the networks, and the image was laterally shifted by 12,60,120 and 270 pixels in the left and right directions. We see that the networks trained on Faces-LFW and ImageNet databases are both able to faithfully reconstruct the phase object at various lateral displacements.

Figures A-11, A-12 show the effect of rotating images on the SLM relative to the upright placement of the images on the SLM on which the neural network was trained at distances z_2 and z_3, respectively. 0 degrees indicates the baseline image placement used during training the networks, and the image was rotated by 90, 180 and 270 degrees. We see that the networks trained on Faces-LFW and ImageNet databases are both able to reconstruct the phase object at varying degrees of rotation.

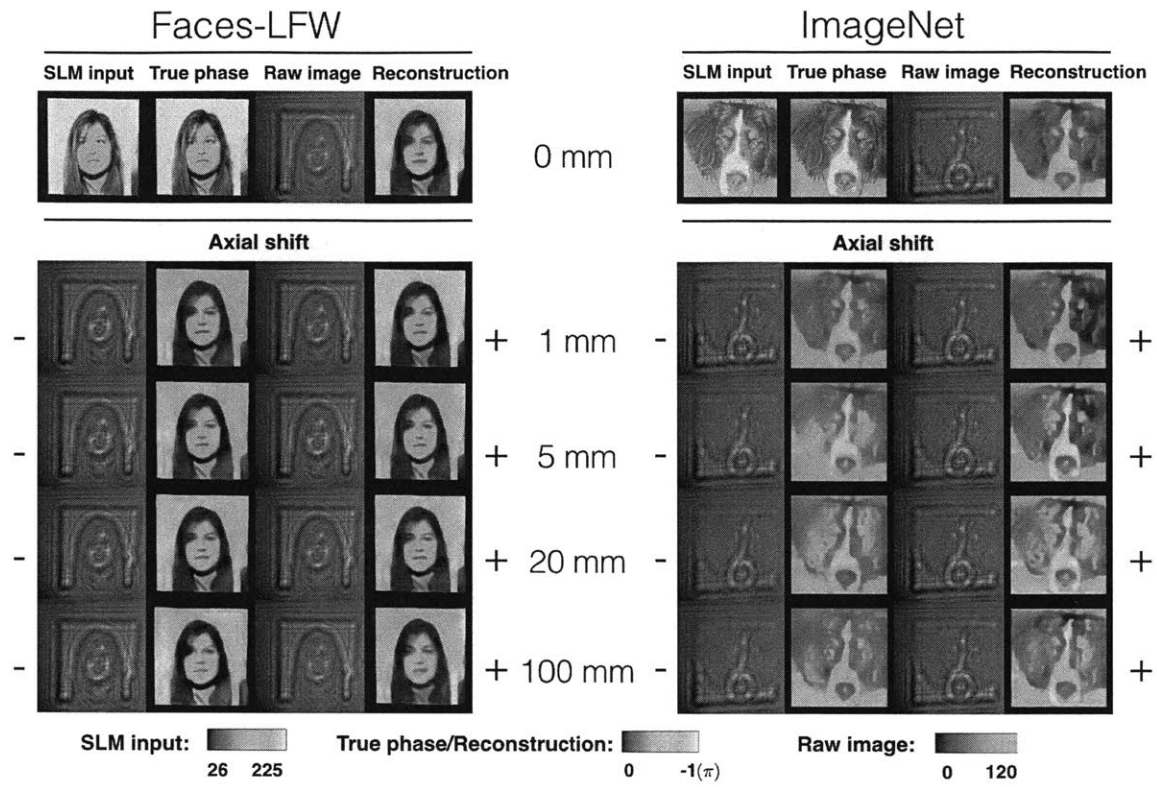

Figure A-8: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to the object-to-sensor distance. The baseline distance on which the network was trained is z_3.

A.4 Linear phase modulation

In the main manuscript and the previous section of the supplementary material, we showed the results of the DNNs which were trained on phase values in the range $[0, -\pi]$. The correspondence between the 8-bit grayscale intensity image projected onto the SLM and the phase values were obtained experimentally as described in the first section, and by least square fitting of 3 piecewise linear segments in the range $[26, 225]$ so as to span $[0, -\pi]$ phase modulation (Fig. A-4). We also trained networks on phase values obtained by least square fitting of a single segment in the range $[30, 224]$ so as to span $[0, -\pi]$ phase modulation (Fig. A-5). This segment naturally results in a poorer fit to the phase modulation in terms of mean square error. Fig. A-13 shows the bar graphs for mean absolute error (MAE) of our trained DNNs at 3 object-to-sensor distances (z_1, z_2 and z_3), as well as the training and test

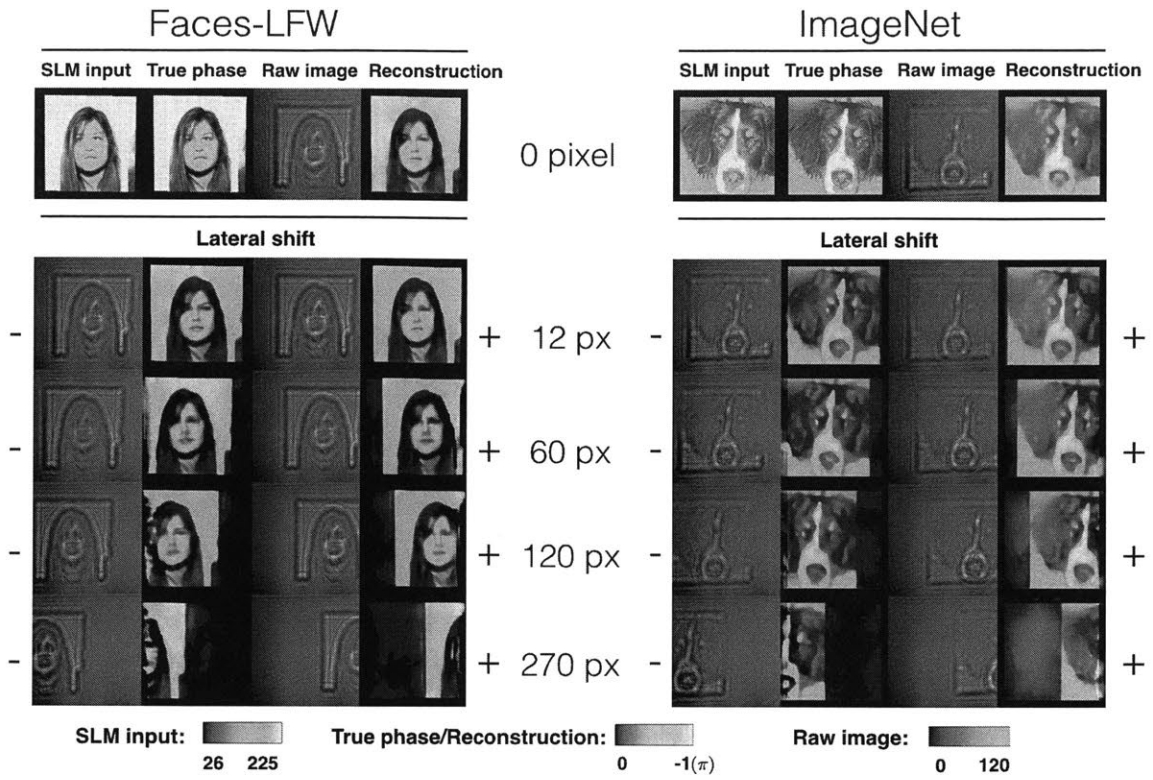

Figure A-9: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to lateral shifts of images on the SLM. The baseline distance on which the network was trained is z_2.

error curves for the network trained on ImageNet database at distance z_1. Fig. A-14 shows qualitative results of our trained DNNs at 3 distances on 7 different datasets, as described in the main manuscript. We see that the mean absolute error (MAE) is comparable to the MAE errors reported in the main manuscript for phase modulation using least square fitting of 3 segments. This suggests that the neural network is fairly robust to the kind of fit we employ for phase modulation.

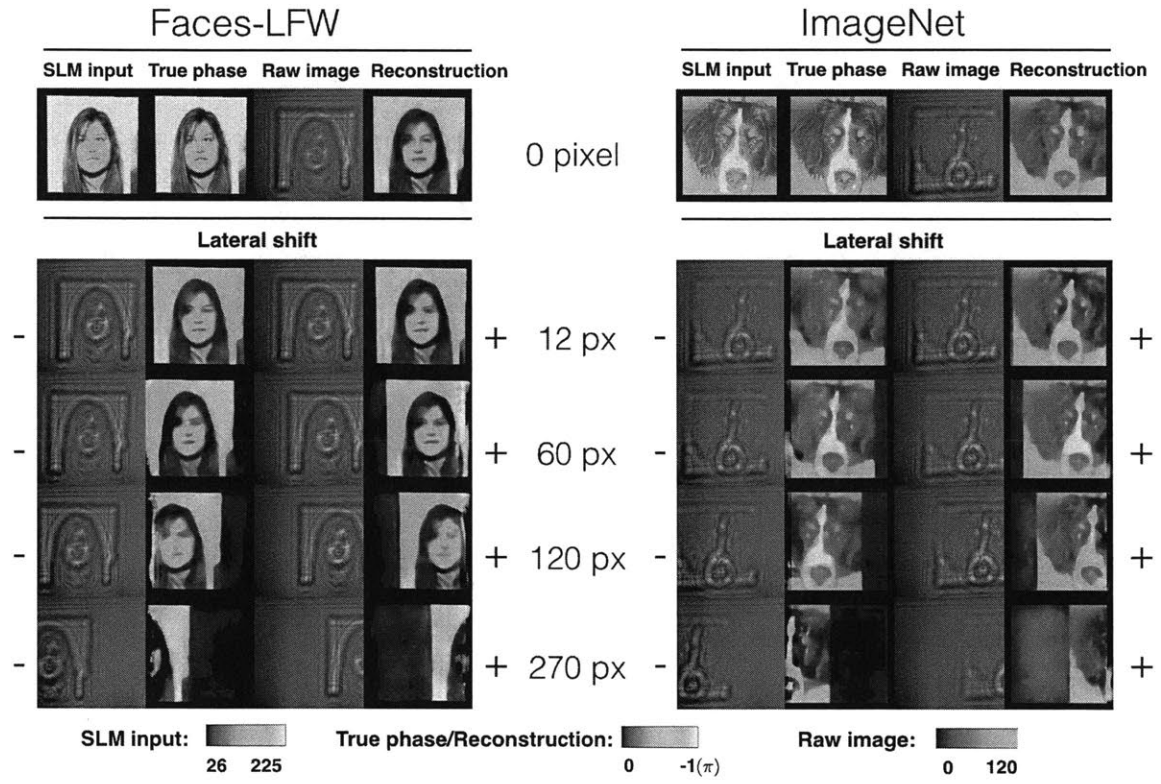

Figure A-10: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to lateral shifts of images on the SLM. The baseline distance on which the network was trained is z_3.

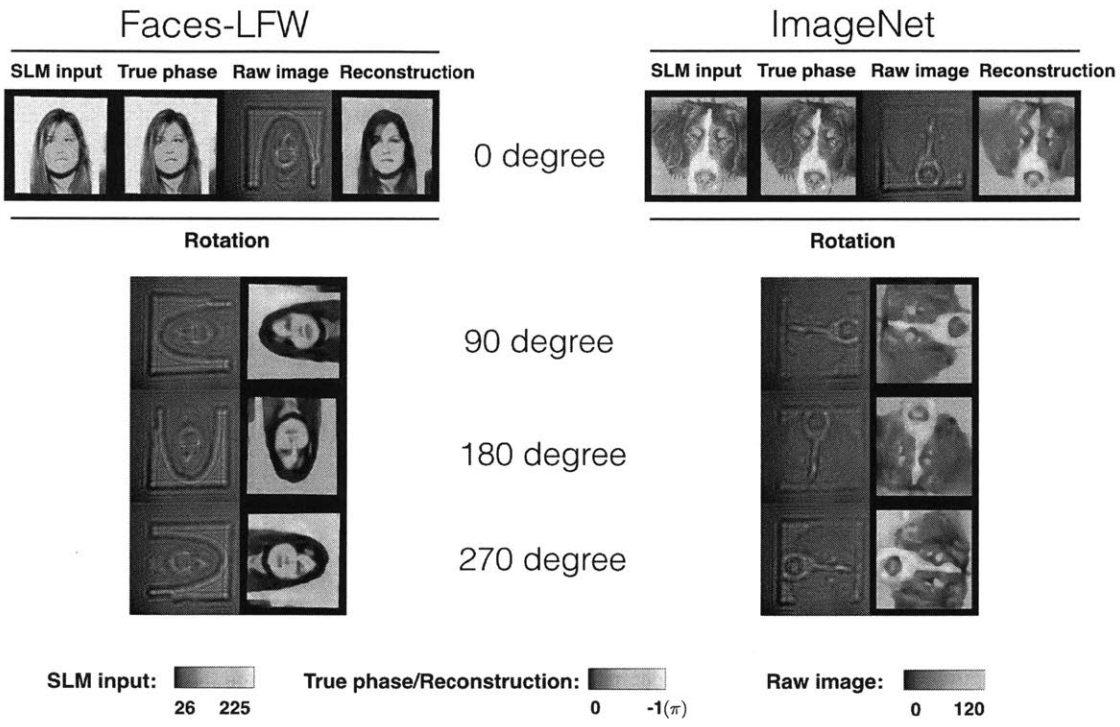

Figure A-11: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to rotation on the SLM. The baseline distance on which the network was trained is z_2.

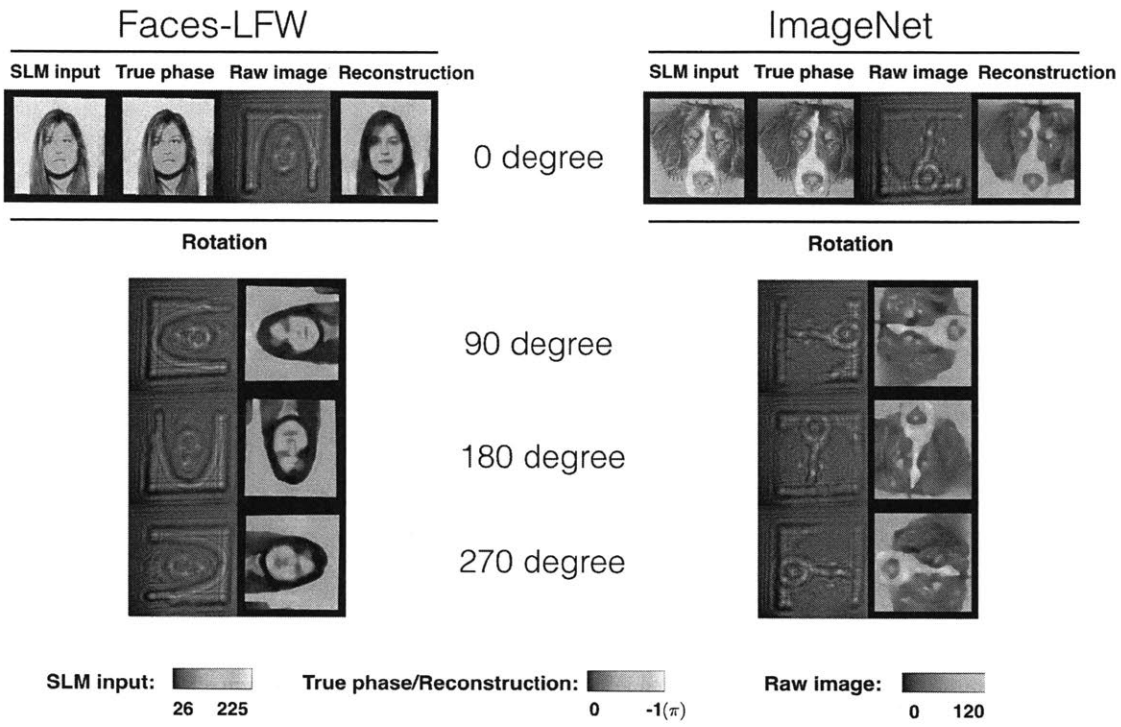

Figure A-12: Qualitative analysis of the sensitivity of the trained deep convolutional neural network to rotation on the SLM. The baseline distance on which the network was trained is z_3.

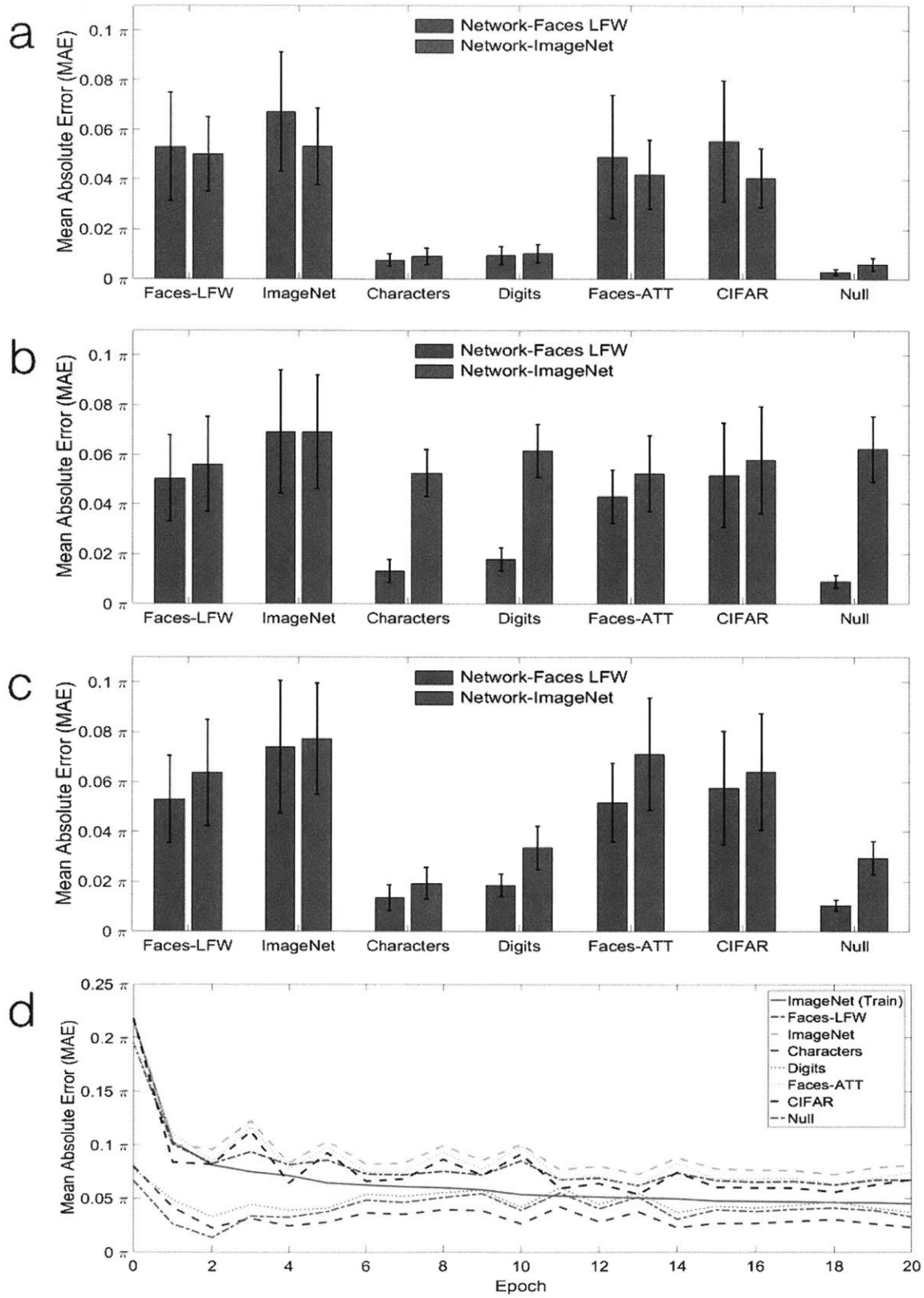

Figure A-13: Quantitative analysis of our trained deep neural networks on phase modulated by a single linear segment for three object-to-sensor distances of (a) z_1, (b) z_2, and (c) z_3 for the DNNs trained on Faces-LFW (blue) and ImageNet (red) on seven datasets. (d) The training and testing error curves for network trained on ImageNet at distance z_1.

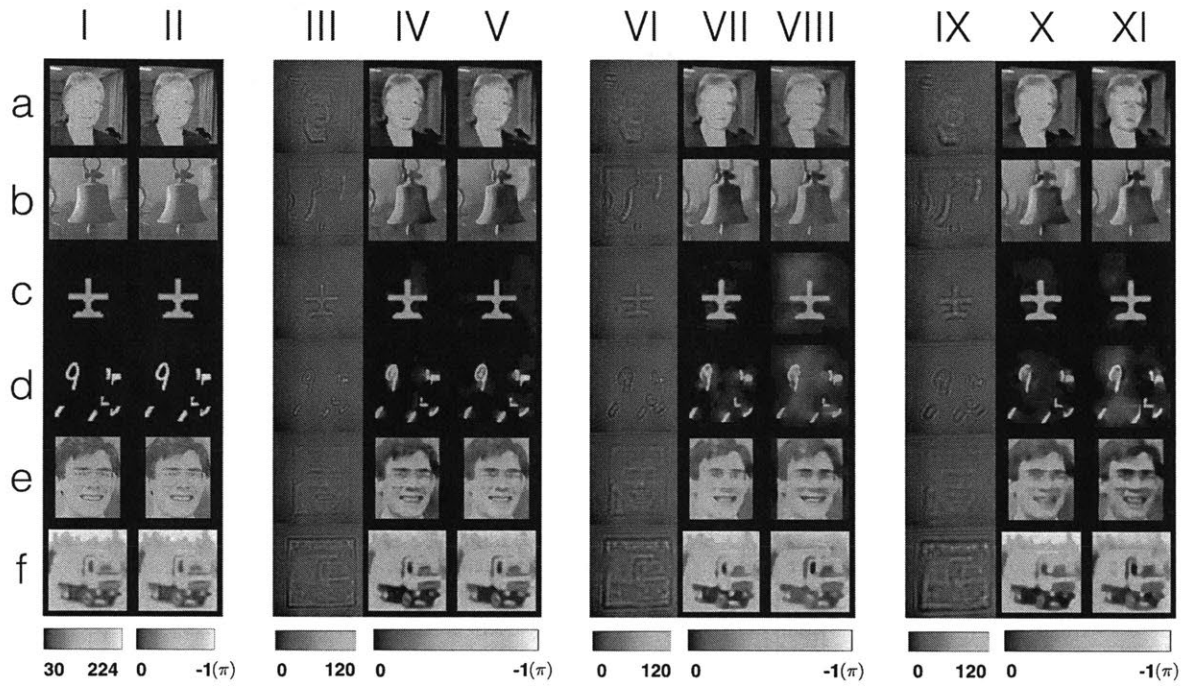

Figure A-14: Qualitative analysis of our trained deep neural networks on phase modulated by a single linear segment for three object-to-sensor distances (z_1, z_2 and z_3) on different datasets. (i,ii) The images in the first two columns are the ground truth inputs to the SLM, and the corresponding phase image calibrated for the SLM; (iii-v) columns show the raw intensity images captured by the CMOS placed at a distance of z_1 which are also inputs to our DNN, the reconstruction by the DNN when trained on Faces-LFW dataset, and the reconstruction by the DNN when trained on ImageNet dataset, respectively. Similarly, for columns (vi-viii) and (ix-xi) the CMOS is placed at a distance of z_2 and z_3, respectively. (a-f) correspond to datasets (a) Faces-LFW, (b) ImageNet, (c) Characters, (d) MNIST Digits, (e) Faces-ATT, and (f) CIFAR, respectively.

Bibliography

[1] Thomas Rüdiger, Heinz Höfler, H. Kreipe, H. Nizze, U. Pfeifer, H. Stein, F. E. Dallenbach, H. Fischer, M. Mengel, and R. Von Wasielewski. "Quality Assurance in Immunohistochemistry Results of an Interlaboratory Trial Involving 172 Pathologists" *Am J Surg Pathol.*, 26(7): 873-882, 2002.

[2] M. E. H. Hammond, D. F. Hayes, M. Dowsett, D. C. Allred, K. L. Hagerty, S. Badve, P. L. Fitzgibbons, G. Francis, N. S. Goldstein, M. Hayes, D. G. Hicks, S. Lester, R. Love, P. B. Mangu, L. McShane, K. Miller, C. K. Osborne, S. Paik, J. Perlmutter, A. Rhodes, H. Sasano, J. N. Schwartz, F. C. G. Sweep, S. Taube, E. E. Torlakovic, P. Valenstein, G. Viale, D. Visscher, T. Wheeler, R. B. Williams, J. L. Wittliff, and A. C. Wolff. "American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version)" *Arch. Pathol. Lab. Med.*, 134(7): e48-72, Jul 2010.

[3] CDC National Program of Cancer Registries "National Program of Cancer Registries (NPCR): United States Cancer Statistics 2010.". https://nccd.cdc.gov/uscs/ (Last Accessed: 13-Jan-2018).

[4] N Howlader, AM Noone, M Krapcho, J Garshell, D Miller, SF Altekruse, CL Kosary, M Yu, J Ruhl, Z Tatalovich, A Mariotto, DR Lewis, HS Chen, and EJ Feuer. "SEER Cancer Statistics Review, 1975-2011, National Cancer Institute, 2014." http://seer.cancer.gov/csr/1975_2011/ (Last Accessed: 13-Jan-2018).

[5] World Health Organization. *World Cancer Report 2014* 2014. Chapter 1.1. ISBN 92-832-0429-8.

[6] Kenneth C Johnson, Anthony B Miller, Neil E Collishaw, Julie R Palmer, S Katharine Hammond, Andrew G Salmon, Kenneth P Cantor, Mark D Miller, Norman F Boyd, John Millar, and Fernand Turcotte. "Active smoking and secondhand smoke increase breast cancer risk: the report of the Canadian Expert Panel on Tobacco Smoke and Breast Cancer Risk (2009)" *Tobacco control*, 20(1): e2

[7] GL Blackburn GL and KA Wang "Dietary fat reduction and breast cancer outcome: results from the Women's Intervention Nutrition Study (WINS)". *The American Journal of Clinical Nutrition.* 86 (3): s878-81.

[8] James Hayes, Ann Richardson, and Chris Frampton "Population attributable risks for modifiable lifestyle factors and breast cancer in New Zealand women". *IMJ.* 43 (11): 1198-1204.

[9] C Kahlenborn, F Modugno, DM Potter, and WB Severs "Oral contraceptive use as a risk factor for premenopausal breast cancer: a meta-analysis". *Mayo Clinic Proceedings.* 81 (10): 1290-302.

[10] XS Wang, ME Armstrong, BJ Cairns, TJ Key, and RC Travis "Shift work and chronic disease: the epidemiological evidence". *Occupational Medicine.* 61 (2): 78-89.

[11] Boris Pasche *Cancer Genetics.* Berlin: Springer. pp. 19-20.

[12] Elizabeth S. McDonald, Amy S. Clark, Julia Tchou, Paul Zhang, and Gary M. Freedman "Clinical Diagnosis and Management of Breast Cancer" *J Nucl Med* 56 (S1): 9S-16S

[13] Breast cancer classification in Wikipedia. Retrieved January 13, 2018, from https://en.wikipedia.org/wiki/Breast_cancer_classification

[14] *National Comprehensive Cancer Network Guidelines for Patients.* Retrieved Jan 13, 2018, from https://www.nccn.org/patients/guidelines/cancers.aspx

[15] A Goldhirsch, E. P. Winer, A S. Coates, R. D. Gelber, M. Piccart-Gebhart, B. Thürlimann, and H.-J. Senn. "Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013." *Ann. Oncol.* 24 (9): 2206-23.Sep. 2013.

[16] A. Goldhirsch, J. N. Ingle, R. D. Gelber, A. S. Coates, B. Thu, and H. Senn. "Thresholds for therapies?: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2009," *Ann. Oncol.* 20(8):1319-29. Aug. 2009.

[17] M. E. Klein, D. J. Dabbs, Y. Shuai, A. M. Brufsky, R. Jankowitz, S. L. Puhalla, and R. Bhargava. "Prediction of the Onco type DX recurrence score?: use of pathology-generated equations derived by linear regression analysis" *Mod. Pathol.* 26(5):658-64. May 2013.

[18] J. Ho, A. V. Parwani, D. M. Jukic, Y. Yagi, L. Anthony, and J. R. Gilbertson. "Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies',' *Hum. Pathol.* 37(3):322-331. Mar. 2006.

[19] J. Gilbertson and Y. Yagi. "Histology, imaging and new diagnostic work-flows in pathology." *Diagn. Pathol.*, 3(S1):S14. Jan. 2008.

[20] R. S. Weinstein, A. R. Graham, L. C. Richter, G. P. Barker, E. a Krupinski, A. M. Lopez, K. a Erps, A. K. Bhattacharyya, Y. Yagi, and J. R. Gilbertson. "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future." *Hum. Pathol.*, 40(8):1057-69. Aug. 2009.

[21] L. Pantanowitz, P. N. Valenstein, A. J. Evans, K. J. Kaplan, J. D. Pfeifer, D. C. Wilbur, L. C. Collins, and T. J. Colgan. "Review of the current state of whole slide imaging in pathology.," *J. Pathol. Inform*, 2(1999):36. Jan. 2011.

[22] Navid Farahani, Anil V Parwani, and Liron Pantanowitz "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives" *Path. and Lab. Med. International* 2015(7):23-33. 2015

[23] Frits Zernike "How I Discovered Phase Contrast" *Science* 121(3141): 345-349, 1955

[24] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21-30, 2008.

[25] E. J. Candès, M. Rudelson, T. Tao and R. Vershynin, "Error correction via linear programming," *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 295-308, 2005.

[26] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, 23, 969-985, 2006.

[27] E. Candès and Y. Plan. "Matrix completion with noise," *P. IEEE* 98(6):925-936 (2010).

[28] Warren McCulloch and Walter Pitts "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics.* 5 (4): 115-133 (1943)

[29] Marvin Minsky and Seymour Papert *Perceptrons: An Introduction to Computational Geometry.* MIT Press. (1969). ISBN 0-262-63022-2.

[30] P.J. Werbos *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences.* Thesis (Ph.D.)-Harvard University, 1974

[31] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner (1998). "Gradient-based learning applied to document recognition" *Proceedings of the IEEE* 86 (11): 2278-2324.

[32] M.J. Bastiaans. "Application of the Wigner distribution function in optics" in *The Wigner Distribution* Elsevier Science, 1997

[33] Lei Tian, Justin Lee, Se Baek Oh, and George Barbastathis. "Experimental compressive phase space tomography," *Opt. Express* 20:8296-8308 (2012)

[34] L. Mandel and E. Wolf. *Optical Coherence and Quantum Optics* Cambridge University Press, 1995

[35] K. Blum, *Density Matrix Theory and Applications* Plenum Press, 1981

[36] K. Itoh and Y. Ohtsuka. "Fourier-transform spectral imaging: retrieval of source information from three-dimensional spatial coherence," *J. Opt. Soc. Am. A* 3:94-100 (1986).

[37] D. L. Marks, R. A. Stack, and D. J. Brady. "Three-dimensional coherence imaging in the Fresnel domain," *Appl. Opt.* 38:1332-1342 (1999).

[38] J. W. Goodman, *Statistical Optics* Wiley-Interscience, 2000

[39] K. A. Nugent, "Wave field determination using three-dimensional intensity information," *Phys. Rev. Lett.* 68:2261-2264 (1992).

[40] M. G. Raymer, M. Beck, and D. McAlister. "Complex wave-field reconstruction using phase-space tomography," *Phys. Rev. Lett.* 72:1137-1140 (1994).

[41] C. Q. Tran, A. G. Peele, A. Roberts, K. A. Nugent, D. Paterson, and I. McNulty, "X-ray imaging: a generalized approach using phase-space tomography," *J. Opt. Soc. Am. A* 22:1691-1700 (2005).

[42] M. Beck, M. G. Raymer, I. A. Walmsley, and V. Wong. "Chronocyclic tomography for measuring the amplitude and phase structure of optical pulses," *Opt. Lett.* 18:2041-2043 (1993).

[43] K. Vogel and H. Risken. "Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase," *Phys. Rev. A* 40:2847-2849 (1989).

[44] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani. "Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum," *Phys. Rev. Lett.* 70:1244-1247 (1993).

[45] U. Leonhardt. "Quantum-state tomography and discrete Wigner function," *Phys. Rev. Lett.* 74:4101-4105 (1995).

[46] C. Kurtsiefer, T. Pfau, and J. Mlynek. "Measurement of the Wigner function of an ensemble of Helium atoms," *Nature* 386:150-153 (1997).

[47] E. Candès, J. Romberg, and T. Tao. "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory* 52:489-509 (2006).

[48] E. Candès, J. Romberg, and T. Tao. "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.* 59:1207-1223 (2006).

[49] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory* 52:1289-1306 (2006).

[50] E. J. Candès and B. Recht. "Exact matrix completion via convex optimization," *Found. Comput. Math.* 9:717-772 (2009).

[51] E. J. Candès and T. Tao. "The power of convex relaxation: near-optimal matrix completion," *IEEE Trans. Inform. Theory* 56:2053-2080 (2010)

[52] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.* 105:150401 (2010).

[53] D. Gross "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inf. Theory* 57:1548-1566 (2011).

[54] E. J. Candès, T. Strohmer, and V. Voroninski. "Phaselift: exact and stable signal recovery from magnitude measurements via convex programming," ArXiv: 1109.4499v1 (2011).

[55] E. J. Candès, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," ArXiv: 1109.0573 (2011).

[56] Y. Shechtman, Y. C. Eldar, A. Szameit, and M. Segev, "Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing," *Opt. Express* 19:14807-14822 (2011).

[57] Emil Wolf. "New theory of partial coherence in the space-frequency domain. Part I: spectra and cross spectra of steady-state sources," *J. Opt. Soc. Am.* 72:343-351 (1982).

[58] D. Pelliccia, A. Y. Nikulin, H. O. Moser, and K. A. Nugent, "Experimental characterization of the coherence properties of hard x-ray sources," *Opt. Express* 19:8073-8078 (2011).

[59] K.-H. Brenner, A. Lohmann, and J. Ojeda-Castaneda, "The ambiguity function as a polar display of the OTF," *Opt. Commun.* 44:323-326 (1983).

[60] K.-H. Brenner and J. Ojeda-Castaneda. "Ambiguity function and Wigner distribution function applied to partially coherent imagery," *Opt. Acta.* 31:213-223 (1984).

[61] J. Tu. "Wave field determination using tomography of the ambiguity function," *Phys. Rev. E* 55:1946-1949 (1997).

[62] R.H.T. Bates and J.M. Rodenburg. "Sub-Angstrom transmission microscopy: a Fourier transform algorithm for microdiffraction plane intensity information," *Ultramicroscopy* 31:303-307 (1989).

[63] J.M. Rodenburg and R.H.T. Bates. "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," *Philos. T. R. Soc. Lond.* 339:521-553 (1992).

[64] B.C. McCallum and J.M. Rodenburg. "2-dimensional optical demonstration of Wigner phase retrieval microscopy in the STEM configuration," *Ultramicroscopy* 45:371-380 (1992).

[65] H.N. Chapman. "Phase-retrieval X-Ray microscopy by Wigner distribution deconvolution," *Ultramicroscopy* 66:153-172 (1996).

[66] J.M. Rodenburg, B.C. McCallum, and P.D. Nellist. "Experimental tests on double resolution coherent imaging via STEM," *Ultramicroscopy* 48:304-314 (1993).

[67] P. Li, T.B. Edo, J.M. Rodenburg. "Ptychographic inversion via Wigner distribution deconvolution: noise suppression and probe design," *Ultramicroscopy* 147:106-113 (2014).

[68] A.M. Maiden, M.J. Humphrey, and J.M. Rodenburg. "Superresolution imaging via ptychography," *J. Opt. Soc. Am.* 28:604-612 (2011).

[69] A.M. Maiden, G.R. Morrison, B. Kaulich, A. Gianocelli, J.M. Rodenburg, "Soft X-ray spectromicroscopy using ptychography with randomly phased illumination," *Nat. Commun.* 4:1669 (2013).

[70] G. Zheng, R. Horstmeyer, C. Yang "Wide-field, high-resolution Fourier ptychographic microscopy," *Nat. Photonics* 7:739-745 (2013).

[71] J. Cai, E. Candès, Z. Shen. "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimiz.* 20(4):1956-1982 (2010).

[72] R.H. Keshavan, A. Montanari and S. Oh. "Matrix completion from noisy entries," arXiv:0906.2027 (2009).

[73] R.H. Keshavan, A. Montanari and S. Oh. "Low-rank matrix completion with noisy observations: a quantitative comparison," arXiv:0910.0921 (2009).

[74] K. Toh and S. Yun. "An accelerated proximal gradient algorithm for nuclear norm regularized least square problems," *Pac. J. Optim.* 6:615-640 (2010).

[75] S. Ma, D. Goldfarb, and L. Chen. "Fixed point and Bregman iterative methods for matrix rank minimization," arXiv:0905.1643 (2009).

[76] K. Lee and Y. Bresler. "Admira: atomic decomposition for minimum rank approximation," arXiv:0905.0044 (2009).

[77] E. Candès and B. Recht. "Exact matrix completion via complex optimization," *Found. Comput. Math* 9:717-772 (2009).

[78] E. Candès, T. Strohmer, and V. Voroninsky. "PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming," arXiv:1109.4499 (2011).

[79] L. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller "Experimental robustness of Fourier ptychography phase retrieval algorithms" *Opt. Express* 23:33214-33240 (2015)

[80] M. Guizar-Sicairos and J. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," *Opt. Express* 16:7264-7278 (2008).

[81] P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," *New J. Phys.* 14:063004 (2012).

[82] J. Cesar da Silva and A. Menzel, "Elementary signals in ptychography," *Opt. Express* 23:33812-33821 (2015)

[83] R. Horstmeyer et al, "Solving ptychography with a convex relaxation," arXiv:1412.1209 (2014)

[84] P. Godard, M. Allain, V. Chamard, and J. Rodenburg, "Noise models for low counting rate coherent diffraction imaging," *Opt. Express* 20:25914-25934 (2012)

[85] Justin Lee and George Barbastathis, "Denoised Wigner distribution deconvolution via low-rank matrix completion," *Opt. Express* 24:20069-20079 (2016)

[86] At&t database of faces. Technical report, AT&T Laboratories Cambridge.

[87] Mario Bertero and Patrizia Boccacci. *Introduction to inverse problems in imaging.* CRC press, 1998.

[88] David J Brady. *Optical imaging and spectroscopy.* John Wiley & Sons, 2009.

[89] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 415-423, 2015.

[90] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, pages 184-199. Springer, 2014.

[91] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197-2202, 2003.

[92] Alexey Dosovitskiy, Jost-Tobias Springenberg, Maxim Tatarchenko, and Thomas Brox. Learning to generate chairs, tables and cars with convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):692-705, 2017.

[93] James R Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*, 3(1):27-29, 1978.

[94] Kunihiko Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological cybernetics*, 20(3-4):121-136, 1975.

[95] RW Gerchberg and WO Saxton. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik (Jena)*, 35:237, 1972.

[96] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[97] Joseph W Goodman and RW Lawrence. Digital image formation from electronically detected holograms. *Applied Physics Letters*, 11(3):77–79, 1967.

[98] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[99] Gao Huang, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten "Densely Connected Convolutional Networks'," *https://arxiv.org/abs/1608.06993*

[100] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554-2558, 1982.

[101] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. Learning-based imaging through scattering media. *Optics Express*, 24(13):13738–13743, 2016.

[102] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst, 2007.

[103] Ju-Seog Jang, Su-Won Jung, Soo-Young Lee, and Sang-Yung Shin. Optical implementation of the Hopfield model for two-dimensional associative memory. *Optics Letters*, 13(3):248-250, 1988.

[104] Hui Ji and Kang Wang. Robust image deblurring with an inaccurate blur kernel. *IEEE Transactions on Image Processing*, 21(4):1624-1634, 2012.

[105] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509-4522, 2017.

[106] YoungJu Jo, Sangjin Park, JaeHwang Jung, Jonghee Yoon, Hosung Joo, Minhyeok Kim, Suk-Jo Jo, Myung-Chul Choi, Sang-Yup Lee, and YongKeun Park. Holographic deep learning for rapid optical screening of anthrax spores. *BioRxiv*, page 109108, 2017.

[107] Ulugbek S Kamilov, Ioannis N Papadopoulos, Morteza H Shoreh, Alexandre Goy, Cedric Vonesch, Michael Unser, and Demetri Psaltis. Learning approach to optical tomography. *Optica*, 2(6):517-522, 2015.

[108] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[109] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097-1105, 2012.

[110] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436-444, 2015.

[111] Yann LeCun et al. Generalization and network design strategies. *Connectionism in perspective*, pages 143-155, 1989.

[112] Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller. Efficient backprop. In *Neural networks: Tricks of the trade*, pages 9-48. Springer, 2012.

[113] Marvin L Minsky and Seymour Papert. *Perceptions: An Introduction to Computational Geomry.* MIT press, 1969.

[114] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529-533, 2015.

[115] SM Popoff, G Lerosey, R Carminati, M Fink, AC Boccara, and S Gigan. Measuring the transmission matrix in optics: an approach to the study and control of light propagation in disordered media. *Physical Review Letters*, 104(10):100601, 2010.

[116] SM Popoff, Geoffroy Lerosey, Mathias Fink, Albert Claude Boccara, and Sylvain Gigan. Image transmission through an opaque material. *ArXiv:1005.0532*, 2010.

[117] Yair Rivenson, Zoltan Gorocs, Harun Gunaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *ArXiv:1705.04709*, 2017.

[118] Yair Rivenson, Yibo Zhang, Harun Gunaydin, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *ArXiv:1705.04286*, 2017.

[119] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[120] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[121] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[122] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *ArXiv:1702.08516*, 2017.

[123] Norbert Streibl. Phase imaging by the transport equation of intensity. *Optics Communications*, 49(1):6-10, 1984.

[124] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 769-777, 2015.

[125] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1-9, 2015.

[126] Michael Reed Teague. Deterministic phase retrieval: a GreenâĂŹs function solution. *JOSA*, 73(11):1434-1441, 1983.

[127] Andrey Nikolayevich Tikhonov. On the stability of inverse problems. In *Dokl. Akad. Nauk SSSR*, volume 39, pages 195-198, 1943.

[128] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 7. MIT press Cambridge, MA, 1949.

[129] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Advances in Neural Information Processing Systems*, pages 341–349, 2012.

[130] Li Xu, Jimmy SJ Ren, Ce Liu, and Jiaya Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*, pages 1790–1798, 2014.

[131] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In *International Conference on Learning Representations*, 2016.

[132] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.

[133] AM Maiden and JM Rodenburg. "An improved ptychographical phase retrieval algorithm for diffractive imaging." *Ultramicroscopy* 109(10):1256-62. Sep 2009

[134] Andrew Maiden, Daniel Johnson, and Peng Li. "Further improvements to the ptychographical iterative engine," *Optica* 4:736-745 (2017)

[135] Xiaoze Ou, Guoan Zheng, and Changhuei Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," *Opt. Express* 22:4960-4972 (2014)

[136] A. M. Maiden, M. J. Humphry, and J. M. Rodenburg. "Ptychographic transmission microscopy in three dimensions using a multi-slice approach," *J. Opt. Soc. Am. A* 29:1606-1614 (2012)

[137] Lei Tian and Laura Waller "3D intensity and phase imaging from light field measurements in an LED array microscope," *Optica* 2:104-111 (2015)

[138] Ludovic Roux, Daniel Racoceanu, Nicolas Lomenie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frederique Capron, Catherine Genestie, Gilles Le Naour, Metin N Gurcan "Mitosis detection in breast cancer histological images An ICPR 2012 contest" *J Pathol Inform*, 4:8 (2013)

[139] Dan C. Ciresan, Alessandro Giusti, Luca M. Gambardella, and Jurgen Schmidhuber. "Mitosis detection in breast cancer histology images with deep neural networks." *International Conference on Medical Image Computing and Computer-assisted Intervention* pp. 411-418. Springer, Berlin, Heidelberg, 2013.

[140] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck. "Deep learning for identifying metastatic breast cancer." arXiv:1606.05718 (2016).

[141] Shuai Li, Mo Deng, Justin Lee, Ayan Sinha, and George Barbastathis "Imaging through glass diffusers using densely connected convolutional networks." arXiv:1711.06810

[142] A. Sinha, O. M' Saad, J. Lee, S. Li, and G. Barbastathis. "Neural Network-based Inversion of the Quantitative Phase Imaging Problem" *Focus on Microscopy* 2017.

[143] Yair Rivenson, Zoltan Gorocs, Harun Gunaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. "Deep learning microscopy," *Optica* 4:1437-1443 (2017)

[144] Shuai Li, Ayan Sinha, Justin Lee, and George Barbastathis. "Quantitative phase microscopy using deep neural networks." *SPIE Photonics West Bios* 2018

[145] Jerome H. Milgram and Weichang Li. "Computational Reconstruction of Images from Holograms," *Appl. Opt.* 41:853-864 (2002)

[146] Wensheng Chen, Lei Tian, Shakil Rehman, Zhengyun Zhang, Heow Pueh Lee, and George Barbastathis. "Empirical concentration bounds for compressive holographic bubble imaging based on a Mie scattering model," *Opt. Express* 23:4715-4725 (2015)

[147] David J. Brady, Kerkil Choi, Daniel L. Marks, Ryoichi Horisaki, and Sehoon Lim. "Compressive Holography," *Opt. Express* 17:13040-13049 (2009)

[148] J. R. Fienup. "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.* 3:27-29 (1978)

[149] Heinz H. Bauschke, Patrick L. Combettes, and D. Russell Luke. "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," *J. Opt. Soc. Am. A* 19:1334-1345 (2002)