Investigating the robustness of a deep learning-based method for quantitative phase retrieval from propagation-based x-ray phase contrast measurements under laboratory conditions

Rucha Deshpande, Ashish Avachat, Frank J Brooks and Mark A Anastasio

1 Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, United States of America
2 Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA, United States of America
3 Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL, United States of America
* Author to whom any correspondence should be addressed.
E-mail: maa@illinois.edu

Keywords: quantitative phase retrieval, propagation-based x-ray phase contrast imaging, laboratory conditions, deep learning-based method

Abstract

Objective. Quantitative phase retrieval (QPR) in propagation-based x-ray phase contrast imaging of heterogeneous and structurally complicated objects is challenging under laboratory conditions due to partial spatial coherence and polychromaticity. A deep learning-based method (DLBM) provides a nonlinear approach to this problem while not being constrained by restrictive assumptions about object properties and beam coherence. The objective of this work is to assess a DLBM for its applicability under practical scenarios by evaluating its robustness and generalizability under typical experimental variations.

Approach. Towards this end, an end-to-end DLBM was employed for QPR under laboratory conditions and its robustness was investigated across various system and object conditions. The robustness of the method was tested via varying propagation distances and its generalizability with respect to object structure and experimental data was also tested.

Main results. Although the end-to-end DLBM was stable under the studied variations, its successful deployment was found to be affected by choices pertaining to data pre-processing, network training considerations and system modeling.

Significance. To our knowledge, we demonstrated for the first time, the potential applicability of an end-to-end learning-based QPR method, trained on simulated data, to experimental propagation-based x-ray phase contrast measurements acquired under laboratory conditions with a commercial x-ray source and a conventional detector. We considered conditions of polychromaticity, partial spatial coherence, and high noise levels, typical to laboratory conditions. This work further explored the robustness of this method to practical variations in propagation distances and object structure with the goal of assessing its potential for experimental use. Such an exploration of any DLBM (irrespective of its network architecture) before practical deployment provides an understanding of its potential behavior under experimental settings.

1. Introduction

X-ray phase-contrast (XPC) imaging enables the acquisition of information about the spatial distribution of the complex valued refractive index of an object, which complements the attenuation contrast provided by conventional x-ray imaging (Snigirev et al 1995). As a result, XPC imaging has been employed in fields ranging from material science (Stevenson et al 2003, Mayo et al 2012) to biomedical science (Bravin et al 2012), where attenuation contrast might be insufficient for resolving certain object features. Furthermore, the advent of polychromatic sources with high spatial coherence further expanded the applicability of XPC imaging to laboratory conditions (Wilkins et al 1996, Hemberg et al 2003, Wilkins et al 2014). Various imaging setups, such as, interferometry-based (Bonse and Hart 1965), grating-based (Pfeiffer 2012), analyzer-based (Chapman et al...
and propagation-based (PB-XPC) (Snigirev et al 1995, Wilkins et al 1996) have been employed for XPC imaging, of which the last is particularly simple in terms of the experimental setup. However, images acquired under laboratory conditions are degraded by partial spatial and temporal coherence, causing some loss in image quality as compared to imaging via synchrotron sources. Thus, any method for the recovery of phase effects from mixed-contrast data acquired in laboratory conditions must compensate for partial coherence effects and be stable under typical system variations.

Several methods for QPR from PB-XPC measurements have been proposed over the last few decades (Langer et al 2008, Burvall et al 2011, Luu et al 2011, Davidoiu et al 2014, Häggmark et al 2017). These methods differ in the quantity estimated, the type of source spectrum, the number of measurements required, assumptions about the object properties, and the linearization scheme used to simplify the governing equations (Paganin and Nugent 1998, Paganin et al 2002, Gureyev et al 2006, Guigay et al 2007, Arhatari et al 2008, Langer et al 2008, Beltran et al 2010, Luu et al 2011, Davidoiu et al 2013, Gürsoy and Das 2013, Häggmark et al 2017, Mohan et al 2020). A learning-based solution to QPR from PB-XPC measurements could be a new approach that allows the circumvention of assumptions about the object, and system linearity. More recently, DLBMs, such as those employing convolutional neural networks (CNNs), have been employed as a solution to the phase retrieval problem in the holographic regime (Zeng et al 2021), far-field regime for ptychographic imaging (Cherukara et al 2020, Harder 2021), and under monochromatic conditions in the near-field regime (Mom et al 2022, Wu et al 2022). However, their applicability to experimental data acquired via PB-XPC imaging under laboratory conditions has not been shown. Most importantly, before employing a DLBM under actual experimental conditions, it is essential to first demonstrate its stability under typical system conditions and material variations. Second, it is equally important to test the generalizability or tolerance of the method towards practical, experimental variations or imprecision. Towards the goal of evaluating the robustness of a DLBM under practical, laboratory conditions, we first employed an end-to-end, DLBM as a nonlinear approach to the QPR problem under partial spatial coherence and polychromaticity conditions; this method is also applicable to complex, multi-material objects manifesting mixed-contrast. We then evaluated the performance of this method under varying system conditions and object conditions and finally, quantified its generalizability towards previously unencountered variations in object structure, system settings, and experimental data.

Thus, the goal of this work was two-fold: an investigation of robustness of the proposed method via simulations, and the exploration of the extent of usefulness or applicability of the method when it was employed naively on experimental data. Demonstrating robustness of the method via stylized simulation studies is an important first step because such demonstrations are much more difficult to perform experimentally. In this work, we do not claim that robustness in our simulation studies guarantees robustness in any experimental study, but we do view passing the simulated robustness tests we describe as being necessary for passing more elaborate experimental tests that are well-beyond the intended scope of this work. Furthermore, the purpose of the experimental study was to investigate the feasibility of employing a DLBM to perform quantitative phase retrieval (QPR) on experimental data. The extent of the applicability of such an approach will depend on the realism of the physical modeling in the simulations, which can certainly be enhanced in future studies.

2. Background

2.1. The phase retrieval problem

The QPR problem for XPC imaging is to estimate the phase induced by the presence of an object in the path of an incident x-ray wavefield from intensity measurements alone. Under laboratory conditions, the presence of partial temporal, and spatial coherence worsens the quality of these measurements. Although propagation-based XPC imaging is robust to imperfect temporal coherence (Wilkins et al 1996, Paganin et al 2006), its tolerance to spatial incoherence is significantly worse (Paganin et al 2006).

In the presented study, we seek to estimate the projected phase that is defined by the imaginary component of the wavefield on the contact plane corresponding to a single energy alone from intensity measurements of a mixed-contrast object, acquired under laboratory conditions (considering a polychromatic source spectrum). Typically, for estimation of the wavefield phase when a mixed-contrast object is considered, at least two measurements that contain complementary information are required (Gureyev and Wilkins 1998, Gureyev et al 2000, Burvall et al 2011). This may be achieved by employing spectral detectors, spectrum switching, multiple monochromatic measurements, or measurements at multiple distances (Gureyev and Wilkins 1998). Alternatively, some solutions relate the attenuation and phase effects via the object thickness (Paganin et al 2002, Beltran et al 2010) to estimate phase from a single measurement. In the present work, two measurements were acquired by varying propagation distances and using a conventional, integrating detector (Gureyev and Wilkins 1998, Paganin and Nugent 1998).
2.2. Canonical measurement model
Consider a thin, multi-material object located at a distance $R_i$ from the source that possesses compact support (which allows for theoretical guarantees of uniqueness in the phase retrieval problem), illuminated by a monochromatic plane wave $U(r, \omega)$ propagating along the optical axis $z$. The object is characterized by its energy-dependent, complex-valued refractive index given by:

$$n(r, \omega, z) = 1 - \delta(r, \omega, z) + i\beta(r, \omega, z),$$

where $r$ represents a position in two-dimensional space corresponding to the plane transverse to the optical axis, $\omega$ represents the temporal frequency, $\delta$ and $\beta$ respectively represent the real and imaginary parts of the complex refractive index, and $i \equiv \sqrt{-1}$.

Variations in the phase and amplitude of the forward ($z > 0$) propagating, monochromatic wavefield $U(r, \omega)$ are induced after its interaction with the object. The phase shift induced by the object can be represented as:

$$\phi(r, \omega) = -k \int \delta(r, \omega, z) dz,$$

where $k = 2\pi/\lambda$ is the wavenumber corresponding to $\omega$. Similarly, the amplitude modulus $M(r, \omega) = \exp[-A(r, \omega)]$, which regulates the attenuation contrast is impacted as:

$$A(r, \omega) = k \int \beta(r, \omega, z) dz.$$

Thus, the transmitted wavefield $U(r, \omega, 0)$ immediately behind the object—that is, at the contact plane where $z = 0$—is given by:

$$U(r, \omega, 0) = \exp[i\phi(r, \omega) - A(r, \omega)] U(r, \omega) = T(r, \omega) U(r, \omega),$$

where $T(r, \omega)$ is the transmission function. The intensity distribution corresponding to the transmitted wavefield represents the contact plane image:

$$I(r, \omega, 0) = |U(r, \omega, 0)|^2.$$  

On propagation through free space, the perturbations in the transmitted wavefield modulate the amplitude and phase of the wavefield. The wavefield at the downstream detector plane ($z = R_2$) can be expressed as:

$$U(r, \omega, R_2) = H(\omega, R_2) * U(r, \omega, 0),$$

where $H$ represents the Fresnel propagator (Paganin et al 2006) and $*$ represents the two-dimensional convolution operation. The corresponding intensity distribution is given by:

$$I(r, \omega, R_2) = |U(r, \omega, R_2)|^2.$$

Considering a polychromatic spectrum and illumination by a spherical wave, the intensity on the downstream detector plane can be expressed as (Gureyev and Wilkins 1998, Paganin et al 2006):

$$I_{\text{poly}}(r/M, R_2/M) = M^2 \int d\omega S(\omega) D(\omega) I(r, \omega, R_2),$$

where $D(\omega)$ is the frequency-dependent efficiency of the detector response, $S(\omega)$ is the spectral weight due to the source spectrum, and $M = (R_1 + R_2)/R_1$ denotes the magnification.

The overall system blur resulting from the source and detector characteristics can be modeled as a convolution with the system point spread function (PSF). Additionally, measurement noise modeled as a mixed Poisson–Gaussian random variable. Thus, the final measured image is given by:

$$I_{\text{poly}}(r/M, R_2/M, \sigma) = M^2 \int d\omega S(\omega) D(\omega) I(r, \omega, R_2) + \text{PSF}(r, \sigma) + \eta,$$

where $\sigma$ denotes the width of the system blur, $\text{PSF}(r, \sigma)$ denotes the system point-spread-function and $\eta$ is the additive system noise.

2.3. Related work
2.3.1. Conventional methods for phase retrieval in PB-XPC
Many popular phase retrieval methods from PB-XPC measurements involve a linear solution to the phase retrieval problem under monochromatic source conditions (Paganin et al 2002, Langer et al 2008, Burvall et al 2011). These methods are typically based on the transport-of-intensity equation (TIE) formulation (Teague 1983) or the contrast transfer function formulation (Pogány et al 1997) and may employ one (Paganin et al 2002, 2020), two (Paganin and Nugent 1998), or more (Guigay et al 2007) measurements. While such deterministic methods are simple to implement and physically intuitive, they require restrictive assumptions such as a single material object, weak absorption, or slowly varying phase. Alternatively, iterative methods—also under monochromatic conditions—have been proposed in the linear formulation (Yan et al 2010, Lu et al 2011) and nonlinear formulations (Davidou et al 2012, 2014, Mohan et al 2020). Phase retrieval from polychromatic measurements, typical to laboratory conditions, has involved (i) extension of deterministic formulations (Gureyev and Wilkins 1998, Lohr et al 2020) that are subject to assumptions about objects or
system settings or (ii) iterative methods, demonstrated on very thin (∼1 µm) objects, and with a linearized system model (Carroll et al. 2017) of the inherently nonlinear mapping between intensity and phase. Thus, there is a need for more widely applicable QPR methods for use with PB-XPC imaging that are suitable for experimental imaging of heterogeneous objects under laboratory conditions and do not rely on assumptions about object properties or system linearity. Towards this goal, DLBMs represent potential solutions that address these issues.

2.3.2. DLBMs for phase retrieval
Prior studies of DLBMs for phase retrieval in optical imaging in the holographic regime have shown promise (Cherukara et al. 2018, Deng et al. 2020, Kang et al. 2020, Zhang et al. 2020, Li et al. 2021a, 2021b, Luo et al. 2021, Wijesinghe and Dholakia 2021, Zeng et al. 2021). Recently, a few methods have also explored their applicability for XPC imaging in the near-field regime (Zhang et al. 2021, Mom et al. 2022, Wu et al. 2022, Xu et al. 2022). However, most DLBMs for phase retrieval from XPC measurements have been proposed for monochromatic conditions. Although two of these recent studies (Wu et al. 2022, Xu et al. 2022) assume a laboratory setup, the first study (Wu et al. 2022) considers a single effective energy for numerical studies, while the second (Xu et al. 2022) involves a grating-based imaging setup. In the present work, we employ and assess a DLBM for (i) QPR from simulated PB-XPC measurements where simulation parameters are consistent with laboratory conditions and with consideration of multi-material objects, (ii) robustness to commonly encountered practical variations in system parameters and object structure and (iii) generalizability to experimental data. Thus, this work aims to investigate the practical applicability and limitations of employing an DLBM for QPR from heterogeneous and structurally complicated objects imaged via PB-XPC under laboratory conditions.

3. Methods

3.1. Overview of approach and study goals
We propose a learning-based solution to the QPR problem under conditions of polychromaticity and partial coherence, typical to laboratory conditions. This approach circumvents the assumptions of the existing methods by learning an analytically intractable, nonlinear mapping from the acquired intensity measurements to the object-induced wavefield phase. A DLBM seeks to learn this mapping from a training dataset constituting sufficient representations or instances of this mapping. Each instance of this mapping consists of two input intensity measurements—one acquired at the contact plane and the second at a downstream detector plane, and a target phase map that represents the wavefield phase on the contact plane corresponding to the peak energy of the spectrum. The rationale for this setup is that at least two inputs are generally required for unambiguous phase retrieval when a multi-material object provides both attenuation, and phase contrast (Gureyev and Wilkins 1998, Gureyev et al. 2000, Burvall et al. 2011). Furthermore, to retrieve phase corresponding to a single energy, from polychromatic data, a network employed as a DLBM is expected to parse the effects of the spectral distribution and learn the mapping between intensity and phase corresponding to only the peak energy. Issues of uniqueness of this mapping are referred to in other works (Nugent 2007, Burvall et al. 2011). In this study we consider typical practical conditions and the absence of phenomena such as phase vortices and zeros.

The proposed DLBM employs an encoder–decoder like U-Net architecture (Ronneberger et al. 2015), which has previously demonstrated a high effectiveness for data-driven solutions for inverse problems (Jin et al. 2017). However, when employing any DLBM, it is not sufficient to merely demonstrate the capability of the method to solve such a problem, but it is also necessary to ensure that the method can be employed stably under variations in system or object conditions that may be practically encountered. Besides stability, considerations of generalizability and dataset sufficiency are also important. Thus, the focus of this work is to investigate the robustness of a learning-based solution for QPR, particularly with regard to practical utility. Specifically, the following properties of a DLBM employed for QPR from PB-XPC measurements under laboratory conditions will be explored: (i) stability under typical laboratory conditions of system blur, noise, propagation distance, and within distribution of the training data, (ii) quantitative accuracy, (iii) qualitative performance related to retention of features of interest, (iv) robustness to practical variations in propagation distances and object structure out of distribution of the training data. We conducted computer-simulations to enable a systematic study of these issues. Additionally, experimental data was utilized to investigate the direct experimental generalizability of the DLBM trained on simulated data. The following sub-sections detail the methodology employed in the studies.

3.2. Numerical phantom design
Two classes of phantoms were employed in the computer-simulation studies: a phantom comprising uniform spheres, and a complex phantom comprising spheres and ellipsoids. The phantoms employed in this work are
The spheres did not intersect each other in 3D although the projected thickness might still show overlaps in 2D as plane to ensure compact support for the object even in the downstream measured image. This also ensured that of optical thickness. The phantoms described below are optically thin. For the designed to be increasingly complicated in terms of object materials, structure, and overlap, but not in the sense variable radii drawn from a beta distribution with parameters—mean: 845 μm, std dev: 20%, range: 422–1690 μm. The physical size of the detector was 13.5 mm × 13.5 mm, and the spheres were placed randomly within the field-of-view (FOV), with at least a distance of 3.6 mm away from the detector edges at the contact plane to ensure compact support for the object even in the downstream measured image. This also ensured that the spheres did not intersect each other in 3D although the projected thickness might still show overlaps in 2D as the spheres were separated only along the projection axis.

The following configurations of the spheres phantoms were considered: (a) single material phantom: soft-tissue in air, (b) multi-material phantom without any embedding medium: four biological tissues (lung, liver, cartilage, intestine), and (c) multi-material phantom including an embedding material: the embedding material corresponds to muscle. The refractive indices of the three embedded materials in the last case were obtained by increasing material density such that the resulting material densities were 1.5, 2 and 2.5 times of the original \((\delta = 4.2 \times 10^{-7} \text{ and } \beta = 2.4 \times 10^{-10} \text{ at } 24 \text{ keV})\). In case (c), the total object thickness was taken as 13.5 mm, which also induced attenuation contrast. This value equals the detector side length and thus, describes the object as a cube. Note that object thickness can be varied and this value was chosen merely to demonstrate the applicability of the method in a regime of realistic samples demonstrating phase contrast (over a centimetre), going beyond very thin or weak attenuation samples (few microns). This phantom with an embedding material was employed to test whether a deep learning-based solution does indeed respect the physical intuition that phase retrieval depends only on the relative differences in the refractive indices of the foreground and the background, while remaining unaffected by the presence of an embedding medium. In other words, the goal of the corresponding study (see section 3.5.2) was to test whether a DLBM has any potential applicability when the phantom contains an embedding medium with thickness of about a centimeter, given the additional difficulties in learning due to the presence of the embedding medium. However, results from simulation studies may not be directly applicable to experimental settings, as additional studies that address beam hardening and other physical factors will need to be conducted.

The complex-valued refractive indices were computed by use of the xraylib library (Schoonjans et al 2011) and elemental compositions of tissues (Russo 2017).

The second class of phantoms, referred to as the ‘complex phantoms’, provided a greater variation in shapes and sizes of the individual components forming an object than the spheres phantoms. This phantom was designed specifically for the assessment of the generalizability to variations in object structure. Each realization of a complex phantom comprised 64 structures drawn uniformly at random from a distribution of four distinct shapes: large spheres (mean radius: 845 μm, std dev.: 0.5 × mean), small spheres (mean radius: 422 μm, std dev.: 0.5 × mean), large ellipsoids (mean semi-axes lengths: 845 μm, 1690 μm, 845 μm, std dev.: 0.5 × mean), and thin ellipsoids (mean semi-axes lengths: 211 μm, 1690 μm, 211 μm, std dev.: 0.05 × mean). Thus, while employing simple, analytically described shapes such as spheres and ellipsoids, this phantom provided greater structural complexity than the spheres phantom due to multiple, overlapping (in projection) constituent structures. The phantom design is not unique to the experiment and thus, the same inferences could be obtained by employing another complex phantom to conduct the relevant experiment. The refractive index allocated to this phantom corresponded to that of soft-tissue \((\delta = 4.09 \times 10^{-7} \text{ and } \beta = 2.2 \times 10^{-10} \text{ at } 24 \text{ keV})\).

### 3.3. System simulation and image formation

For use in network training, simulated PB-XPC data were produced by imaging the phantoms virtually. The source spectrum was modeled to match an x-ray source with a liquid-metal-jet anode (MetalJet D2, Excillum) operated at 70 kVp, at 110 W power (peak energy of 24 keV and equivalent energy 19 keV), and filtered through 0.9 mm Aluminium, which resulted in a change in equivalent energy from 19 keV to 32 keV post filtering. This reduced low-energy photons while retaining the intensity peaks in the source spectrum. The spectrum was resampled in steps of 1 keV in the Bremsstrahlung or non-peak region, and 0.2 keV near the peak, yielding 65 discrete energy levels in total. The detector was modeled as a CsI(Tl) scintillator-based x-ray imager with a pixel pitch of about 13 μm. At each of these energies, the virtual image was formed by the interaction of the object with the incident radiation, modeled as a convolution of the object transmission function with the Fresnel propagator (Voelz 2011) in the Fourier domain (refer to section 2.1). To simulate the effects of a cone beam geometry, the Fresnel scaling theorem (Paganin et al 2006) was employed after propagation. For the geometries simulated in this work, the source-to-object distance \((R_s)\) was set to 1 m and the object-to-detector distance \((R_d)\) was varied from 0.25 m to 1 m, yielding magnifications in the range 1.25–2. All simulations were oversampled by a factor of 2 to ensure sufficient sampling (Voelz 2011). The monochromatic images at each discrete energy bin were then combined to form a single polychromatic image according to their spectral weights and the energy-specific
detective sensitivity. Next, the effect of system blur was modeled as a convolution of the virtual image with a Gaussian kernel whose width was determined by the full-width-half-maximum (FWHM) of the system blur, ranging from 39 μm to 92 μm, and corresponding to 3, 5 and 7 detector pixels respectively. Last, the image was downsampled to yield a 1024 × 1024 image and system noise was added, modeled as a mixed Poisson–Gaussian random variable possessing zero mean and standard deviation ranging from 0.5% to 5% of individual pixel intensity. A contact plane image (I₀) and the downstream measured image (Iₙ) form a single input pair to the U-Net, and examples are shown in figure 1. The corresponding target image that the U-Net seeks to approximate was specified as a phase map of the projected thickness at 24 keV energy, and computed as \( \phi_{E=24\text{keV}} = -k_0\delta_{E=24\text{keV}}t \) where \( t \) is the projected thickness of the object. Achieving similar statistics between the contact image and the downstream image was implied in the simulations and the problem setup. Each input image, contact and downstream, is assumed to be appropriately corrected by its respective image and the downstream image was implied in the simulations and the problem setup. Each input image, contact and downstream, is appropriate and required.

**Figure 1.** Sample input intensity images under various system conditions. (Top row) Contact plane images. (Bottom row) Intensity images acquired after propagation. This image-pair serves as the two-channel input to a DLBM. Case 1: low blur (FWHM: 39 μm) and low noise (0.5%). Case 2: high blur (FWHM: 92 μm) and high noise (5%). Case 3: low \( R_z \) (0.25 m), high noise (5%). Case 4: high \( R_z \) (1.00 m), high noise (5%).

3.4. Network trainings

A U-Net deep neural network architecture (Ronneberger et al 2015) was employed that contained two input- and one output-channel. All images were of dimension 1024 × 1024. It should be noted that the FOV for each of the two input images: \( I_0, I_z \), is different, and determined by the magnification. The two input channels were normalized independently over the entire dataset. This ensured that the mapping from intensity to phase is retained while also aiding training stability. As observed in our numerical experiments, while global normalization ensures the retention of the information of this physical mapping, a per-image normalization breaks this mapping and renders the problem unsolvable. Per-sample normalization maps the range of intensity values seen in one sample to the corresponding phase values. As the extremes of intensity of each sample are different, per-slice normalization could result in a non-unique mapping between inputs and outputs after normalization, and thus, potentially make phase retrieval impossible. Thus, global normalization, as opposed to per-sample normalization, is appropriate and required.

A network of depth three was chosen and trained to minimize the mean squared error (MSE) loss using an Adam optimizer (learning rate = \( 5 \times 10^{-3} \) to \( 5 \times 10^{-6} \), \( \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \times 10^{-7} \)). Other parameter choices included: kernels of size 3 × 3 in the encoder/decoder blocks, upsampling via nearest-neighbor interpolation, downsampling via max-pooling, and skip connections that copied feature maps from the encoding arm to the decoding arm. Model training for the case of objects without an embedding medium (zero background phase value) was performed for 100 epochs with a batch size of 1, and the model with the least validation loss was chosen for inference. For the case of objects embedded within a surrounding medium (non-zero background phase value), training was performed with an adaptive learning rate schedule (learning rate = \( 5 \times 10^{-3} \) to \( 5 \times 10^{-5} \)) over 160 epochs. The training, validation and test sets consisted of 2000, 1000 and 1000 images respectively. Model training was performed on Nvidia GeForce GTX TitanX, 1080-Ti and A100 GPUs.
and typically took 15 to 60 hours per training. The inference time for a single sample is 16 s and for a test set of 1000 samples is about 12 min on a Nvidia TitanX GPU.

For the numerical studies of variations in system conditions, materials and phantoms, a separate network was trained for each combination of system and object parameters. For the generalizability studies, a single network trained for a certain object and geometry combination was tested on multiple, out-of-distribution datasets.

3.5. Description of experiments
The performance of the DLBM was assessed under variations in system parameters: blur and noise—varied jointly, object-to-detector distance ($R_2$), and object composition: multi-material objects in the presence or absence of an embedding medium. Here onward, these cases are referred to as SYS-BN, SYS-PROP, and MAT-EMB, MAT-NEMB respectively. Next, generalization studies were performed to investigate the robustness of a pre-trained DLBM when applied to data that differed in some characteristics from the data employed in model training. In the machine learning literature, such data are referred to as ‘out-of-distribution’ data (OOD). In this work, the generalizability of the method was tested on simulated OOD datasets corresponding to variations in propagation distances or $R_2$, and complexity of object structures. The two studies are referred to as OOD-PROP and OOD-OBJ respectively. Subsequently, experimental data were also employed as a type of OOD data; this study is referred to as OOD-EXPT.

Three measures of performance were used for assessment in the simulation studies described above: % normalized root-mean-square error over the object (NRMSE), structural similarity index measure (SSIM) (Wang et al 2004) and peak signal-to-noise ratio (PSNR) over the entire image. The % NRMSE is defined as:

$$\text{%NRMSE} = \frac{\| \phi_{\text{pred}} - \phi_{\text{true}} \|_2}{\| \phi_{\text{true}} \|_2} \times 100.$$  

All comparisons were made against two popular analytic methods for single material objects by Paganin et al (Paganin et al 2020), or multi-material objects embedded in a medium by Beltran et al (Beltran et al 2010), here onward referred to as PG and BT respectively. Each of these methods was implemented at two energies: the peak energy of the spectrum, corresponding to the energy of the target phase map for the DLBM, and the equivalent energy of the spectrum. The two variations of the PG/BT methods are referred to as PG/BT-peak and PG/BT-eqvt respectively. These methods are stable under high noise conditions, and have been known to provide greater accuracy and low-frequency stability than the two-plane TIE method (Paganin and Nugent 1998, Paganin et al 2002). Hence, they are chosen to provide higher benchmarks for comparison. As both methods, PG and BT, recover the object thickness, the phase map was obtained by employing the known values for $k\delta (E_{\text{eqv}} = 24keV)$. Furthermore, the multi-material method BT (Beltran et al 2010) requires knowledge of object locations, and this was supplied to the method.

Descriptions of all experiments are given in the subsections below and a summary is provided in table 1.

3.5.1. Behavior of the method under various system conditions
All phantoms in SYS-BN and SYS-PROP consisted of a single material, with the refractive index values representative of soft-tissue. This ensured that the impact of system variations was not confounded by material variations, and allowed a fair comparison with conventional methods intended for use with homogeneous objects. For the SYS-BN study, blur and noise were varied jointly in the range 39 μm—92 μm FWHM and 0.5%...
—5% respectively. For the SYS-PROP study, \( R_2 \) was varied from 0.25 to 1 m. For both studies, the other system parameters were held constant as specified in table 1.

3.5.2. Behavior of the method under variations in materials and phantoms
As the distribution of refractive indices varies with materials and energy, it is important to assess the extent to which the DLBM can learn how to recover the phase when the object is heterogeneous. Hence, to study the performance of the DLBM in the presence of material heterogeneity, two cases were considered. The first case: MAT-NEMB, consisted of multiple materials with no embedding medium, and the second case: MAT-EMB, consisted of multiple materials in an embedding medium. These correspond to configurations (b) and (c) of the spheres phantom respectively (see section 3.2). System settings were fixed as specified in table 1. In the two cases, results from the DLBM were compared against the conventional method PG for MAT-NEMB, and BT for MAT-EMB, employed separately for each material, followed by splicing.

3.5.3. Robustness to propagation distances
For the OOD-PROP study, a DLBM trained with data corresponding to a fixed set of system parameters: \( R_1 = 1 \) m, \( R_2 = 0.5 \) m, blur = 66 \( \mu \)m FWHM and noise = 5%, was employed to perform phase retrieval by use of OOD test datasets produced by introducing variations in propagation distances in the range: \( R_2 = 0.4-0.6 \) m, except \( R_2 = 0.5 \) m, which was the baseline case. As the flat-field intensity is also impacted by variations in the propagation distance, additional error (besides error due to OOD measurements) may be incurred due to incorrect flat-field normalization. Hence, two cases of normalization were considered for the OOD measurements \( I_{z=\text{OOD}} \), where the subscript indicates that the propagation distance \( z \) was OOD and did not correspond to training measurements \( I_{z=0.5\mu m} \). The two cases of normalization were: (i) Unscaled \( I_z \), where \( I_{z=\text{OOD}} \) measurements were flat-field corrected by the use of flat-field measured at \( I_{z=0.5\mu m} \), and (ii) Scaled \( I_z \), where \( I_{z=\text{OOD}} \) measurements were flat-field corrected appropriately by the use of corresponding flat-field (acquired at the same OOD value of \( z \)).

3.5.4. Robustness to object structure
To test the robustness to variations in object structure (OOD-OBJ), a DLBM trained on the complex phantom (described in section 3.2) was employed for phase retrieval on test data that corresponded to structurally distinct objects imaged under the same system settings as the training data (refer to table 1), and for the same material allocations (object material: soft-tissue). The test data consisted of two examples: an irregular structure consisting of multiple, vessel-like sub-structures (Beutel et al. 2000) with mean radii: 40, 8 pixels, and a regular structure constituting of equidistant ellipsoids with semi-axes lengths: 20, 200, 50 pixels, placed on a grid.

3.5.5. Generalizability to experimental data
For the OOD-EXPT study, polystyrene spheres of radius 1.94 ± 0.05 mm were experimentally imaged to acquire two images: a contact plane image (\( R_1 = 3.16 \) m, \( R_2 = 0.02 \) m) and a downstream measured image (\( R_1 = 1.24 \) m, \( R_2 = 1.94 \) m, FOV = 20 mm). The focal spot size was 12 \( \mu \)m, while all other system parameters were as described before (in section 3.3 and table 1). The OOD test dataset comprised 15 measurements, each of which corresponded to an image-pair acquired by imaging a physical phantom consisting of 5 polystyrene spheres. The 15 measurements were all acquired under the same experimental settings to allow for the effects of experimental variations from the beam and to provide multiple instances of system noise. A network trained on simulated data corresponding to the spheres phantom, but matched in material, distribution of radii, and system geometry, was employed for predictions on experimental data. Prior to prediction, the experimental data were flat-field corrected, intensity outliers were removed by saturating pixels in the extreme 0.5 percentiles, and the data were normalized as per the training dataset. After phase retrieval, accuracy in maximum thickness was computed as:

\[
\text{%Accuracy} = \left[1 - \frac{\text{true} - \text{predicted}}{\text{true}}\right] \times 100. \tag{10}
\]

4. Results

4.1. Behavior of the network under various system conditions
Qualitative results for the studies pertaining to variations in the physical parameters: SYS-BN and SYS-PROP, are presented in figure 2. Generally, the features of the object were well formed although some blur at the edges can be observed in cases with high noise levels (5%). Predictions from the DLBM have noticeably lower noise and, thus, stronger bias as compared to results from the conventional methods under high noise conditions.
However, some high-frequency, residual artifacts were observed in the difference images for the DLBM predictions. Additionally, very low amplitude artifacts, spatially corresponding to object locations at the magnified FOV were also present (see figure 3), potentially resulting from the joint tasks of QPR and demagnification.
Quantitative results, reported as mean and standard deviation computed over a test set of 1000 samples, from the first study: SYS-BN, wherein noise and system blur were varied to study the combined effects of these physical factors on the performance of the DLBM demonstrate that although the NRMSE for the DLBM was < 10% in all cases, increasing system noise and system blur levels resulted in degraded quantitative performance (see figure 4) as expected. However, the NRMSE was more impacted—6.5% compared to 0.7%—by an increase in system noise from 0.5% to 5% than by increase in system blur from 39 to 92 μm, at intermediate values of the other parameter. Similar trends were observed for the SSIM and PSNR (see figure 4), indicating the superior quantitative performance of the DLBM, especially at high levels of noise. This revealed that, although blur and noise worsened the performance of the DLBM, it was stable within typical ranges of these system parameters found in laboratory conditions, and has better quantitative performance than conventional methods even at high levels of noise under the simulated conditions.

Similarly, within SYS-PROP, to study the impact of varying the object-to-detector distance, three choices of $R_2$: 0.25m, 0.5 m and 1 m ($R_1 = 1$ m) were considered, corresponding to magnifications of 1.25, 1.5 and 2 respectively. It was observed that the change in the quantitative performance (NRMSE) was negligible for the DLBM across all $R_2$ values, and the NRMSE was consistently lower than the corresponding values of the conventional methods (figure 5). Note that the effects of magnification and propagation distances are related, and the DLBM is expected to account for both effects simultaneously. Moreover, the impact of variation in noise (0.5% to 5%) on quantitative performance of the DLBM was greater (≈6.5% compared to ≈0.3%) than the impact of variations in $R_2$ (from 0.25 m to 1 m).

Figure 3. Examples of typical artifacts seen in phase predictions. (Left) Sample difference image shows two types of artifacts: high-frequency artifacts at the object edges, and low-amplitude, residual artifacts corresponding to the object locations in the magnified input image. (Right) Cropped section from the left image shows that the residual error is negligible (< 1 radian) as compared to the maximum absolute phase in the ground truth (≈86 radians). Additionally, it was observed that the edge artifacts are not uniform in all directions implying that rotationally invariant post-processing techniques for artifact removal may not suffice.

Figure 4. Results for SYS-BN: variations in blur and noise, over a test set of 1000 samples (mean and standard deviation). Shaded error bands are shown for the intermediate blur case, i.e. blur = 66 μm FWHM. The DLBM outperforms conventional methods in all three metrics, particularly PSNR and SSIM at high noise levels. This indicates better retention of qualitative features and denoising effects due to the DLBM as compared to the conventional methods. Furthermore, an increase in noise levels from 0.5% to 5% had a much larger effect on performance than variation in blur in the range 39–92 μm FWHM, as seen in the near overlap of values of all performance metrics on varying blur at each noise level.
4.2. Behavior of the network under variations in materials and phantoms

Qualitative results from the multi-material phantom study (figure 6) demonstrate that although phase predictions from all methods exhibited well-formed features, the DLBM clearly outperformed the conventional methods in regions of overlapping structures and in terms of denoising performance. Quantitative results from both cases of the multi-material phantom study show that the DLBM outperforms PG and BT (Beltran et al 2010) in terms of pixel-wise accuracy (NRMSE) and PSNR (see table 2). Although the SSIM was similar ($\approx 0.96$) for the DLBM and the conventional methods, quantitative errors were clearly greater for the conventional methods than the DLBM as seen in the NRMSE ($>30\%$ in MAT-NEMB and $>0.2\%$ in MAT-EMB).
MAT-EMB and PSNR (>20 dB). Note that the NRMSE appeared to be much lower in the presence of an embedding medium, i.e. MAT-EMB. This is because it is a measure of relative error, and contains the effects of a larger background phase value (≈680 radians) as compared to that in MAT-NEMB (0 radians). Quantitative and qualitative results from the multi-material phantom, together indicate that the DLBM could successfully perform QPR by recovering a nonlinear mapping between phase and intensity, and not simply a linear mapping for a single ‘effective’ material.

Finally, the effect of foreground area was studied to determine whether increasing the information content per realization, i.e. whether a greater number of pixels having a value other than that of the constant background within a realization aided quantitative performance. This is important to understand because the MSE loss function was employed for training. The system settings were: $R_1 = 1\text{ m}$, $R_2 = 0.5\text{ m}$, blur = 66 $\mu\text{m}$ FWHM and noise = 5%. It was found that foreground area variations in the range 10%–30% did not cause any significant change in the quantitative performance of the network (<0.5%), indicating that the training ensemble of images produced from the designed phantoms contained sufficient instances of intensity-phase mappings to learn QPR.

### 4.3. Generalization studies

#### 4.3.1. Robustness to propagation distances

As seen in figure 7, the pre-trained network was robust for variations of at least 20% in the propagation distance when $I_2$ was scaled according to Case (ii) with negligible change in quantitative performance. This was different from Case (i), which resulted in an increase of up to 7.5% in NRMSE for the same range of $R_2$. For this baseline model in particular, it was also observed that the SSIM for some of the unscaled datasets, close to the baseline case, was slightly higher than the baseline. As SSIM is highly sensitive to normalization, some variation in this metric may be observed around the baseline value, and across different baseline models trained on the same data. However, the trends in performance over increasing variation in $R_2$ remain the same. NRMSE and PSNR have been observed as being more robust across multiple models in this regard. Qualitatively, although features were well formed, a slight asymmetry was also observed occasionally; this might be due to the mismatch in magnification between the training and testing datasets. Thus, this numerical study suggests that the DLBM,

![Figure 7. Generalization performance on varying propagation distance (OOD-PROP) for a network trained with data corresponding to $R_1 = 0.5\text{ m}$ and tested on datasets with $R_2$ ranging from 0.4 to 0.6 m. Scaling $I_2$ by the correct flat-field, i.e. corresponding to $R_2$(test) improves the robustness of the network for all three metrics as compared to the unscaled dataset, i.e. corresponding to $R_2$(train). This effect is more pronounced at lower values of $R_2$.](image)

Table 2. Quantitative results for multi-material phantoms in the absence of an embedding medium: MAT-NEMB, and in its presence: MAT-EMB. For a multi-material phantom, irrespective of the presence or absence of an embedding medium, the DLBM clearly outperforms PG and BT in terms of NRMSE and PSNR, while demonstrating performance similar to PG/BT methods in terms of the SSIM.

| Multi-material object cases | Measure of quality | Methods                  |  |
|----------------------------|-------------------|--------------------------|
|                            |                   | DLBM | PG/BT-equiv | PG/BT-peak |
| Embedding medium absent    | NRMSE (%)         | 9 ± 1 | 40 ± 13 | 37 ± 11 |
|                            | SSIM              | 0.98 | 0.95 | 0.96 |
|                            | PSNR (dB)         | 41 ± 2 | 26 ± 2 | 26 ± 2 |
| Embedding medium present   | NRMSE (%)         | 0.5 ± 0.04 | 0.7 ± 0.2 | 0.9 ± 0.4 |
|                            | SSIM              | 0.99 | 0.96 | 0.96 |
|                            | PSNR (dB)         | 41 ± 2 | 27 ± 1 | 25 ± 3 |
with proper normalization of data, might have sufficient robustness for application under experimental settings, where the accuracy of the measured propagation distance might be inexact due to practical considerations.

4.3.2. Robustness to object structure
It was observed that the DLBM generalized well to variations in object structure (figure 8), based on qualitative results as well as the relative (not absolute) error in predictions, as quantified by the NRMSE. In both cases, the NRMSE was < 10%—at par with the within-distribution case. This revealed that a DLBM is not constrained by accurate structural modeling of the test object. This, in turn, implies that the DLBM could be used for inference from complex object measurements even when trained with measurements produced by imaging a collection of structurally simpler objects.

4.3.3. Generalizability to experimental data
Quantitative phase predictions produced by the DLBM yielded superior accuracy in maximum thickness (94 ± 6%), as opposed to predictions from PG-eqvt (66 ± 1%) and PG-peak (86 ± 1%). The DLBM performed comparably in terms of equivalent diameter (97 ± 2% compared to 99 ± 1% for both PG methods) and roundness (90 ± 1% compared to 90 ± 7% for both PG methods). Note that these results for the DLBM indicate a baseline performance in the absence of any transfer learning, which may potentially improve the generalizability to experimental data by addressing certain inaccuracies in system modeling.

Qualitative results (mean and standard deviation images over 15 predictions) are shown in figure 9. The spheres appear well-formed and noise is largely suppressed. The background does show instances of slightly inadequate noise suppression, possibly due to the mismatch in the simulated noise distribution and the experimental system noise; however, noise suppression is far superior to the PG methods. Thus, the DLBM outperformed PG in terms of noise suppression and estimation of maximum depth, while performing comparably in terms of qualitative aspects as measured by equivalent diameter and roundness. This indicates that an DLBM trained on simulated data can generalize well to experimental data when system modeling is accurate.
5. Discussion

The significance of the present work is that it is a demonstration that the DLBM could be an effective solution to QPR in PB-XPC imaging under laboratory conditions. In addition, we provide a framework of studies that is important in understanding the potential behavior of any learning-based method. Although such a framework may not be required for analytical methods, whose behavior may be predicted based on a formula, understanding the behavior of learning-based methods requires an empirical exploration. Here, we do not claim that all learning-based methods would be successful at QPR irrespective of architecture and training choices, but that a deep learning-based approach, as demonstrated via one architecture, and one system, could be applicable to the QPR problem. Thus, inferences drawn in the context of numerical and experimental studies employing this one method, referred to as the DLBM, might vary for a different method. The behavior of another learning-based method also could be explored via the framework described in this work.

Furthermore, observations from simulated studies of stability and robustness of the DLBM may vary under practical, experimental conditions due to the effects of physical parameters not considered in simulations. Therefore, additional experimental studies are required in this direction before the general translation of any DLBM to experimental conditions.

5.1. Behavior under system and object variations

The DLBM is stable in quantitative performance under various simulated conditions of system blur, noise and object-to-detector distances as seen in figures 4 and 5. In addition, the qualitative comparison shows superior denoising as compared to the conventional methods. However, a comprehensive study of relative denoising properties of the foreground and background is beyond the scope of this work. Similar studies exploring the relative denoising properties and their relation with factors such as variations in pixel counts between the foreground and background, or relative magnitudes of material properties, would provide more insight towards the experimental translation of the DLBM. Together, the stability in quantitative and qualitative performance over typical system settings under simulated conditions suggests that an DLBM may be applicable over a wide range of typical system conditions, and with fewer limiting assumptions compared to popular alternatives. Most importantly, results from the generalization studies demonstrate the robustness of the DLBM to minor variations in propagation distances typical to practical conditions, where the precision of acquiring images at a given propagation distance might be low due to limitations of instrumentation. Furthermore, the robustness to changes in object structure can prove beneficial in the absence of accurate models of object or large experimental datasets. Last, results from the experimental study demonstrated that the DLBM could perform reasonably well on data acquired from an experimental phantom, indicating that such an approach could hold merit in experimental settings, provided that the system modeling is accurate. However, for a general translation of the

![Figure 9. Mean and standard deviation computed over 15 predictions of phase from experimental images of polystyrene spheres (OOD-EXPT) as predicted by the DLBM (left), PG-eqvt (center) and PG-peak (right). Qualitative results from experimental data demonstrate generally well formed spheres in the mean prediction for all methods and quantitatively superior denoising for the DLBM in the standard deviation image. The true value of phase at the center of each projected sphere is $-96 \pm 2$ radians. Despite predicting well-formed structures, PG-eqvt and PG-peak are observed to underestimate the true value of phase, unlike the DLBM, which achieves greater accuracy.](image-url)
to a variety of experimental conditions, more practical studies specific to experimental factors such as the effects of scatter, beam inhomogeneity, and various embedding materials might be required. Further improvement in experimental studies potentially could be achieved via explicit modeling of additional system parameters or implicitly via transfer learning methods. By transfer learning methods, we refer to strategies developed specifically to improve the generalizability of deep learning-based methods, when a distribution or domain shift may be encountered by a network at inference time. In the context of this work, such methods may involve training with a limited amount of experimental data after a network is trained on simulated data—thus allowing the network to learn the effects of some experimental parameters that may be absent in the simulations.

5.2. Advantages and disadvantages of a DLBM

A DLBM enables the learning of a nonlinear mapping without strong restrictions on the composition of the object, unlike many analytic methods for QPR (Langer et al 2008). In addition, quantitative phase maps for multi-material objects can be retrieved at once, without pair-wise consideration of materials (Haggmark et al 2017) or post-hoc splicing. However, the accuracy of such an end-to-end DLBM is data dependent: it is affected by the accuracy in system modeling as well as by sufficiency in representation of the underlying mapping of the heterogeneous object to the phase map. Here, representational sufficiency is with regard to material overlaps only and not object structures. Thus, a DLBM trained on a phantom consisting of specific materials could translate well to a structurally distinct phantom consisting of the same materials, as demonstrated in the object structure generalization study. There remains a need to compare end-to-end DLBMs, as employed in this study, against other DLBMs that incorporate system knowledge to investigate improvements in generalizability of a learning-based approach.

5.3. Plausibility of a CNN-based learning method as a solution for QPR

Phase contrast in intensity images manifests as the fringes present at the material edges in an object. The characteristic size of each fringe depends on the following system considerations: system blur, propagation distance, and wavelength, besides object attenuation (Nesterets et al 2005, Gureyev et al 2008). Thus, in order to retrieve phase from polychromatic data under laboratory conditions, the concurrent performance of two operations is expected: (i) implicit filtering or extraction of the dominant, spatio-temporally coherent mode (Zysk et al 2010, Born and Wolf 2013) from the input intensity measurements and, (ii) phase retrieval corresponding to this extracted dominant mode. The first problem may be represented as a deconvolution and denoising problem; the chosen DLBM architecture has been employed as a solution for the two problems in other imaging scenarios (Park et al 2018, Reymann et al 2019, Lee et al 2020). The second problem, QPR from the dominant coherent mode, is an ill-posed, inverse problem. Solution to this problem, in the context of the TIE, requires knowledge of the transverse derivatives (gradient and Laplacian) of the two input intensity images. These operations, too, can be performed by a CNN-based architecture, such as that employed in the DLBM. Thus, it is feasible that a CNN-based method can learn a solution to the polychromatic phase retrieval problem.

5.4. Training considerations

A DLBM aims to learn the mapping between phase and intensity measurements, independent of the phantom structure and design. In order to retain this mapping during data pre-processing, while simultaneously ensuring an appropriate range of data for gradient computation, it is essential to normalize over the entire stack of training images and not over individual slices. Per-slice normalization would render the problem unsolvable. Furthermore, the validation and test datasets are also mapped onto the same, normalized scale. Thus, retention of this inverse mapping has to be ensured not just during pre-processing and training, but also during inference.

As the pixel-wise MSE loss was employed for training, quantitative performance was prioritized over qualitative feature information. Alternatively, if the goal is excellent qualitative performance alone, alternative loss functions that take into account the correlations in object structure such as normalized correlation coefficient (NCC) (Goy et al 2018), perceptual loss (Deng et al 2020) or even CNR, could be employed to focus on qualitative feature information. To this end, a variation in network architecture such as mixed-scale dense networks (Pelt and Sethian 2018, Mom et al 2022) instead of convolutional networks could be employed. Another effect of employing the MSE loss is the strong denoising performance, given the large proportion of pixels forming the background. However, the superior denoising performance might be partially offset by slightly worse performance for low phase contrast-inducing materials present within multi-material objects, and under high noise conditions, as the network training minimizes MSE and prioritizes superior performance on high phase contrast-inducing materials. An alternative training strategy might include material-specific weights in the loss function, or involve training specifically at various quantitative ranges of phase. Furthermore, with a focus on QPR and employment of a pixel-wise loss, high-frequency artifacts were observed at the edges, especially as residues at the locations of the magnified objects. These artifacts were relatively small, and could be
removed via post-processing. Before the practical deployment of any DLBM, it is important to identify, \textit{a priori},
the kind of characteristic artifacts that may be produced by the DLBM as opposed to artifacts from conventional
methods that can be determined given their analytical formulation (Rodgers et al 2020). This could inform a
post-processing method that might be employed to eliminate the same.

Last, in case of multi-material objects, the presence of an embedding material adds additional difficulty in
the context of a deep learning-based solution, in terms of both attenuation, and phase effects. Since the thickness
of the embedding material, and hence attenuation due to the same, varies spatially, the network is expected to
learn and eliminate this effect while parsing the effects of attenuation and phase due to the foreground materials
alone. Furthermore, the constant background phase and phase variations relative to the background could span
different orders of magnitude. With MSE as the loss function, learning occurs at two scales—(i) estimating the
background, which often consists of more pixels than the foreground and thus, could provide large
improvement in MSE if correctly estimated, and (ii) estimating the variations in phase with respect to the
background. In order to improve training performance in this scenario, employing an adaptive learning rate, or
offsetting the constant background value, were both effective strategies.

6. Conclusion

In conclusion, we employed a DLBM and assessed its robustness as a solution for QPR from polychromatic
intensity measurements for PB-XPC imaging under laboratory conditions. It was found that the method is
stable under typical variations in system settings and can be applicable for complex, multi-material objects. Last,
the method generalized well to previously unseen, complex phantoms as well as to minor variations in
propagation distances. The experimental applicability of a DLBM is demonstrated via one architecture and one
system; changes in architecture or major changes in the system would require retraining, followed by
reevaluation of the trained model—similar to the studies conducted in this work. Thus, a DLBM has potential
applicability for QPR from measurements of complex objects imaged experimentally via PB-XPC under
laboratory conditions.

Acknowledgments

This work made use of the Illinois Campus Cluster, a computing resource that is operated by the Illinois Campus
Cluster Program (ICCP) in conjunction with the National Center for Supercomputing Applications (NCSA)
and which is supported by funds from the University of Illinois Urbana-Champaign. This research is also part of
the Delta research computing project, which is supported by the National Science Foundation (award OCI
2005572), and the State of Illinois. Delta is a joint effort of the University of Illinois Urbana-Champaign and its
National Center for Supercomputing Applications. This work was supported in part by NIH awards R01
EB018525, R01 EB020604 and EB031772 (subproject 6366). RD acknowledges support from the Imaging
Sciences Pathway Trainee Fellowship (T32EB01485505). We thank Varun A Kelkar for sharing data employed
for the simulation of test example (i) described in section 4.3.2.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/
10.7910/DVN/ITHRKN. Data will be available from 18 January 2023.

ORCID iDs

Rucha Deshpande  \(\text{https://orcid.org/0000-0002-9984-8647}\)
Ashish Avachat  \(\text{https://orcid.org/0000-0002-7266-3992}\)
Mark A Anastasio  \(\text{https://orcid.org/0000-0002-3192-4172}\)

References

Arhatari B D, Hannah K, Balaur E and Peele A 2008 Phase imaging using a polychromatic x-ray laboratory source \textit{Opt. Express} 16 19950–6
Beltran M A, Paganin D, Uesugi K and Kitchen M J 2010 2d and 3d x-ray phase retrieval of multi-material objects using a single defocus
distance \textit{Opt. Express} 18 6423–36
Deshpande R, Avachat A, Brooks F J and Anastasio M A 2023 Simulated propagation-based x-ray phase contrast (PB-XPC) measurements
under laboratory conditions for the design and evaluation of learning-based methods for quantitative phase retrieval (https://doi.
org/10.7910/DVN/ITHRKN)
Beutel J, Kundel H L, Kim Y, Van Metter R L and Horii S C 2000 Handbook of Medical Imaging (Bellingham, Washington, USA: SPIE Press) vol 3
Bonse U and Hart M 1965 An x-ray interferometer Appl. Phys. Lett. 6 155–6
Born M and Wolf E 2013 Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light (Amsterdam: Elsevier)
Bravin A, Coan P and Suortti T 2012 X-ray phase-contrast imaging: from pre-clinical applications towards clinics Phys. Med. Biol. 58 R1–R33
Burvall A, Lundström U, Takman P A, Larsson D H and Hertz H M 2011 Phase retrieval in x-ray phase-contrast imaging suitable for tomography Opt. Express 19 10359–76
Carroll A J, Van Riessen G A, Balaur E, Dolbnya I P, Tran G N and Peele A G 2017 An iterative method for near-field fresnel region polychromatic phase contrast imaging J. Opt. 19 075003
Chapman D, Thomlinson W, Johnston R, Washburn D, Pisanu E, Gmür N, Zhong Z, Menk R, Arfelli P and Sayers D 1997 Diffraction enhanced x-ray imaging Phys. Med. Biol. 42 2015–23
Cherukuri M J, Nashed Y and Harder R J 2018 Real-time coherent diffraction inversion using deep generative networks Sci. Rep. 8 1–8
Cherukuri M J, Zhou T, Nashed Y, Enfedaque P, Hexemer A, Harder R J and Holt M V 2020 AI-enabled high-resolution scanning coherent diffraction imaging Appl. Phys. Lett. 117 044103
Davidou V, Sibux B, Langer M and Peyrin F 2012 Nonlinear phase retrieval using projection operator and iterative wavelet thresholding IEEE Signal Process Lett. 19 579–82
Davidou V, Sibux B, Langer M and Peyrin F 2013 Nonlinear approaches for the single-distance phase retrieval problem involving regularizations with sparsity constraints Appl. Opt. 52 3977–86
Davidou v et al 2014 Non-linear phase tomography based on fréchet derivative Adv. Comput. Tomogr. 3 39–50
Deng M, Goy A, Li S, Arthur K, Li S and Barbastathis G 2018 Low photon count phase retrieval using deep learning Phys. Rev. Lett. 121 243902
Guigay J P, Langer M, Boistel R and Cloetens P 2007 Mixed transfer function and transport of intensity approach for phase retrieval in the Fresnel region Appl. Phys. Lett. 117 044103
Gureyev T E, Nesterets Y I, Paganin D, Pogany A and Wilkins S 2006 Linear algorithms for phase retrieval in the Fresnel region: II. Partially coherent illumination Opt. Commun. 259 569–80
Gureyev T E, Nesterets Y I, Stevenson A W, Miller P R, Pogany A and Wilkins S W 2008 Some simple rules for contrast, signal-to-noise and resolution in in-line x-ray phase-contrast imaging Opt. Express 16 3232–41
Gürsoy D and Das M 2013 Single-step absorption and phase retrieval with polychromatic x rays using a spectral detector Opt. Lett. 38 1461–3
Häggmark I, Vågberg W, Hertz H M and Burvall A 2017 Comparison of quantitative multi-material phase-retrieval algorithms in propagation-based phase-contrast x-ray tomography Opt. Express 25 33543–58
Harder R 2013 Deep neural networks in real-time coherent diffraction imaging IUCrJ 8 1–3
Hemberg O, Otendal M and Hertz H 2003 Liquid-metal-jet anode electron-impact x-ray source Appl. Phys. Lett. 83 1483–5
Jin K H, McCann M T, Froustey E and Unser M 2017 Deep convolutional neural network for inverse problems in imaging IEEE Trans. Image Process. 26 4509–22
Kang I, Zhang F and Barbastathis G 2020 Phase extraction neural network (phen) with coherent modulation imaging (cmi) for phase retrieval at low photon counts Opt. Express 28 21578–600
Langer M, Cloetens P, Guigay J P and Peyrin F 2008 Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography Med. Phys. 35 4556–66
Lee S, Negishi M, Urakubo H, Kasai H and Ishii S 2020 Mu-net: multi-scale u-net for two-photon microscopy image denoising and restoration Neural Netw. 125 92–105
Li A-C, Chen Y-J, Huang H-M and Luo Y 2021a U-net based isocentre based phase retrieval with quantitative differential phase contrast microscopy Frontiers in Optics + Laser Science 2021, (Technical Digest Series) (Optica Publishing Group) pp W7A–19
Li A-C, Vyas S, Lin Y-H, Huang H-M and Luo Y 2021b Patch-based u-net model for isotropic quantitative differential phase contrast imaging IEEE Trans. Med. Imaging 40 5229–37
Loh R L, Scott C C, Pil-Ali A and Karim K S 2020 A comparison of phase retrieval methods for propagation-based contrast x-ray imaging with polychromatic sources Medical Imaging 2020: Physics of Medical Imaging 11312 (SPIE)
Luo F, Wang J, Zeng J, Zhang L, Zhang B, Xu K and Luo X 2021a Cascaded complex u-net model to solve inverse scattering problems with phaseless-data in the complex domain IEEE Trans. Antennas Propag. 70 6170–60
Luo M et al 2011 Quantitative phase imaging with x-ray source Opt. Express 19 8127–34
Mohan K, Stevenson A W and Wilkins S W 2012 In-line phase-contrast x-ray imaging and tomography for materials science Materials 5 937–65
Mohan K A, Parkinson D Y and Cuadra J A 2020 Constrained non-linear phase retrieval for single distance x-ray phase tomography Electron. Magn. 2020 146–1–8
Mom K, Sibux B and Langer M 2022 Mixed scale dense convolutional networks for x-ray phase contrast imaging Appl. Opt. 61 2497–505
Nesterets Y I, Wilkins S, Gureyev T, Pogany A and Stevenson A 2005 On the optimization of experimental parameters for x-ray in-line phase-contrast imaging Rev. Sci. Instrum. 76 093706
Nugent K 2007 X-ray noninterferometric phase imaging: a unified picture J. Opt. Soc. Am. A 24 536–47
Park J, Hwang D, Kim K Y, Kang S K, Kim Y K and Lee J S 2018 Computed tomography super-resolution using deep convolutional neural network Phys. Med. Biol. 63 145011
Pelt D M and Sethian J A 2018 A mixed-scale dense convolutional neural network for image analysis Proc. Natl. Acad. Sci. 115 254–9
Pfeiffer F 2012 Milestones and basic principles of grating-based x-ray and neutron phase-contrast imaging AIP Conf. Proc. vol 1466 (American Institute of Physics) pp 2–11
Pogany A, Gao D and Wilkins S 1997 Contrast and resolution in imaging with a microfocus x-ray source Rev. Sci. Instrum. 68 2774–82
Reymann M P, Würfl T, Ritt P, Stimpel B, Cachovan M, Vija A H and Maier A 2019 U-net for spect image denoising 2019 IEEE Nuclear Science Symp. and Medical Imaging Conf. (NSS/MIC) (IEEE) pp 1–2
Rodgers G, Schulz G, Deyhle H, Kuo W, Rau C, Weitkamp T and Müller B 2020 Optimizing contrast and spatial resolution in hard x-ray tomography of medically relevant tissues Appl. Phys. Lett. 116 023702
Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Springer) pp 234–41
Russo P 2017 Handbook of X-ray Imaging: Physics and Technology (Boca Raton, FL: CRC Press)
Schoonnans T, Brunetti A, Golosio B, del Rio M S, Solé V A, Ferrero C and Vincze L 2011 The xraylib library for x-ray-matter interactions recent developments Spectrochim. Acta B 66 776–84
Snigirev A, Snigireva I, Kohn V, Kuznetsov S and Schelokov I 1995 On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation Rev. Sci. Instrum. 66 5486–92
Stevenson A W et al 2003 Phase-contrast x-ray imaging with synchrotron radiation for materials science applications Nucl. Instrum. Methods Phys. Res. B 199 427–35
Teague M R 1983 Deterministic phase retrieval: a green’s function solution J. Opt. Soc. Am. 73 1434–41
Voelz D G 2011 Computational Fourier Optics: a MATLAB Tutorial vol 534 (Bellingham, Washington: SPIE Press)
Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity IEEE Trans. Image Process. 13 600–12
Wijesinghe P and Dholakia K 2021 Emergent physics-informed design of deep learning for microscopy J. Phys.: Photon. 3 021003
Wilkins S, Gureyev T E, Gao D, Pogany A and Stevenson A 1996 Phase-contrast imaging using polychromatic hard x-rays Nature 384 355–8
Wilkins S, Nesterets Y I, Gureyev T, Mayo S, Pogany A and Stevenson A 2014 On the evolution and relative merits of hard x-ray phase-contrast imaging methods Phil. Trans. R. Soc. A 372 20130021
Wu Y, Zhang L, Guo S, Zhang L, Gao F, Jia M and Zhou Z 2022 Enhanced phase retrieval via deep concatenation networks for in-line x-ray phase contrast imaging Phys. Med. 95 41–9
Xu Y, Tao S, Bian Y, Bai L, Tian Z, Hao X, Kuang C and Liu X 2022 Single-shot grating-based x-ray phase contrast imaging via generative adversarial network Opt. Lasers Eng. 152 106960
Yan A, Wu X and Liu H 2010 Performance analysis of the attenuation-partition based iterative phase retrieval algorithm for in-line phase-contrast imaging Opt. Express 18 16074–89
Zeng T, Zhu Y and Lam E Y 2021 Deep learning for digital holography: a review Opt. Express 29 40572–93
Zhang Y, Noack M A, Vagovic P, Fezzaa K, Garcia-Moreno F, Ritschel T and Villanueva-Perez P 2021 Phasegan: a deep-learning phase-retrieval approach for unpaired datasets Opt. Express 29 19593–604
Zhang Z, Zheng Y, Xu T, Upadhya A, Lim Y J, Mathews A, Xie L and Lee W M 2020 Holo-unet: hologram-to-hologram neural network restoration for high fidelity low light quantitative phase imaging of live cells Biomed. Opt. Express 11 5478–87
Zysk A M, Schoonover R W, Carney P S and Anastasio M A 2010 Transport of intensity and spectrum for partially coherent fields Opt. Lett. 35 2239–41