



# Reliable deep-learning-based phase imaging with uncertainty quantification

YUJIA XUE, SHIYI CHENG, YUNZHE LI, AND LEI TIAN\*

Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, USA

\*Corresponding author: leitian@bu.edu

Received 7 January 2019; revised 6 April 2019; accepted 7 April 2019 (Doc. ID 357144); published 7 May 2019

Emerging deep-learning (DL)-based techniques have significant potential to revolutionize biomedical imaging. However, one outstanding challenge is the lack of reliability assessment in the DL predictions, whose errors are commonly revealed only in hindsight. Here, we propose a new Bayesian convolutional neural network (BNN)-based framework that overcomes this issue by quantifying the uncertainty of DL predictions. Foremost, we show that BNN-predicted uncertainty maps provide surrogate estimates of the true error from the network model and measurement itself. The uncertainty maps characterize imperfections often unknown in real-world applications, such as noise, model error, incomplete training data, and out-of-distribution testing data. Quantifying this uncertainty provides a per-pixel estimate of the confidence level of the DL prediction as well as the quality of the model and data set. We demonstrate this framework in the application of large space-bandwidth product phase imaging using a physics-guided coded illumination scheme. From only five multiplexed illumination measurements, our BNN predicts gigapixel phase images in both static and dynamic biological samples with quantitative credibility assessment. Furthermore, we show that low-certainty regions can identify spatially and temporally rare biological phenomena. We believe our uncertainty learning framework is widely applicable to many DL-based biomedical imaging techniques for assessing the reliability of DL predictions.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

<https://doi.org/10.1364/OPTICA.6.000618>

## 1. INTRODUCTION

The imaging throughput of traditional techniques is fundamentally limited by the intrinsic trade-off among field of view (FOV), resolution, and acquisition speed. It is well known that the space-bandwidth product (SBP) of an optical system is invariant under any linear canonical transform [1,2]. Further considering super-resolution-type techniques that require multiple measurements, the acquisition time scales linearly with the expanded bandwidth in a single dimension, and quadratically for 2D isotropic resolution enhancement [3,4]. The same scaling law also applies to scanning-based systems for enlarging the FOV. Accordingly, the 3D trade-space spanned by the FOV, resolution, and acquisition speed can be visualized as shown in Fig. 1(a), with a hyperplane defining the achievable imaging attributes, which highlights the linear trade-off among them (for a 1D problem). The imaging techniques of our interest belong to the classical phase-retrieval problem. Despite the extra complexity from the intensity-only, nonlinear measurements, the general scaling law for the achievable imaging attributes follows the same trade-space, as studied both theoretically [5] and experimentally [6,7]. Our first goal here is to investigate the feasibility of bypassing the classical limit imposed by the linear trade-space by combining non-conventional multiplexed measurement schemes and deep learning (DL). By doing so, our technique will open up an expanded design space

that allows a combination of FOV, resolution, and acquisition speed beyond those achievable using conventional phase-retrieval techniques [as illustrated in Fig. 1(a)].

Our work is inspired by the recent demonstration of several DL-based phase-retrieval techniques [8–17], which can be categorized into two classes. The first class focuses on solving the phase-retrieval problem alone using a convolutional neural network (CNN)—no modification to the measurement procedure is made [8–13]. As a result, these techniques generally do not improve the imaging throughput. Nevertheless, using the CNN-based algorithm has been reported to have several benefits, including robustness to noise, scattering, and experimental errors [8–13]. The second class focuses on introducing the physical model into the construction of the CNN. This is done by modeling the image formation process as the initial layers of the CNN [14–17]. As a result, training the CNN jointly optimizes the physical parameters used in the acquisition alongside its computational parameters. However, the effectiveness of this approach relies on the accurate modeling of the image formation process [14], which can be difficult in practice due to the presence of uncalibrated aberrations and other experimental imperfections.

Differing from these two classes, we propose to solve the large-SBP phase-retrieval problem using a *physics-guided* DL approach, which consists of two complementary components. The first

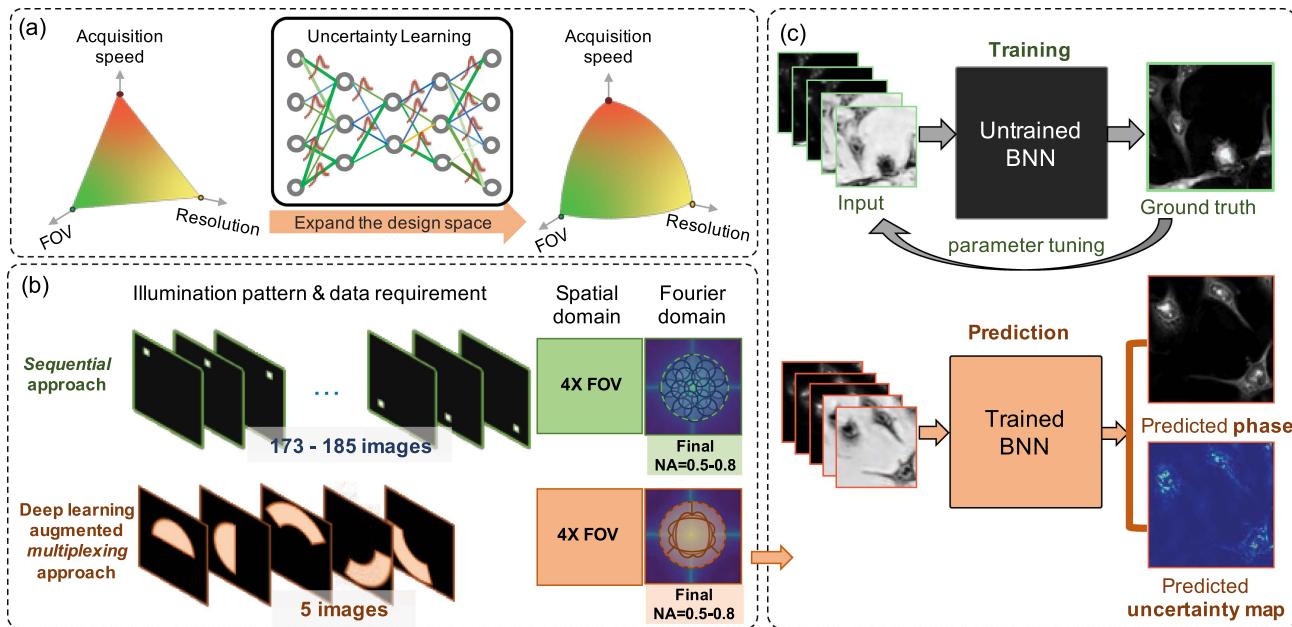
component is a highly measurement-efficient illumination multiplexing strategy designed by two physical principles. First, we exploit asymmetric illumination to encode the phase information into the intensity measurements based on the principle of differential phase contrast (DPC) [18]. Second, we enhance the resolution following the principles of the synthetic aperture [19] and Fourier ptychographic microscopy (FPM) [6] by using oblique illumination to introduce into the measurements high-frequency information that are beyond the native passband of the objective lens. Most importantly, our method uses only *five* coded measurements regardless of the final resolution [Fig. 1(b)], making our technique highly flexible and scalable for large-SBP phase-retrieval problems. As a result, our proposed technique avoids the need to quadratically increase the number of measurements to achieve a higher resolution—a limitation that is imposed by conventional FPM techniques. The reason behind preventing such multiplexed measurements to be used previously is the severe ill-posedness of the resulting inverse problem [7,20–22]. This results in undesirable phase artifacts in the reconstruction from existing multiplexed FPM (mFPM) algorithms. The second component uses DL to overcome the ill-posedness of the inverse problem and complements the new measurement strategy. Specifically, we show that our DL algorithm robustly inverts the physical model and recovers large-SBP phase information from highly multiplexed nonlinear measurements, which would otherwise not be possible.

An important feature of our DL technique is the ability to quantitatively assess its reliability. In particular, we aim to address a common criticism on DL that the error of the prediction cannot be easily evaluated unless the ground truth is known. To address this issue, we develop an uncertainty learning (UL) framework based on the Bayesian convolutional neural network (BNN) [23] [Fig. 1(c)]. We show that the reliability of the BNN prediction can be quantified by two predictive uncertainties, the model uncertainty and the data uncertainty, akin to the epistemic and

aleatoric uncertainties, respectively, in Bayesian analysis [24]. In particular, we show that the model uncertainty allows us to characterize the robustness of our physics-guided DL technique. By training and testing on an ensemble of CNNs, the BNN quantifies the variabilities intrinsic to the model without “cherry-picking” the results [23]. In addition, we show that the data uncertainty allows assessing the randomness of the predictions that originate from data imperfections [23], including noise, incompleteness in the training data, and error due to out-of-distribution testing data.

In order to rigorously quantify the reliability of the BNN predictions, an important step is to perform statistical data analysis. We develop a procedure to relate the BNN output to Bayesian statistical metrics, including credibility, credible interval, and reliability diagram. By doing so, our work establishes a comprehensive procedure for evaluating the reliability of our DL-based phase-retrieval technique.

By capturing experimental data on two different computational microscopy platforms, we justify our proposition that our technique is applicable to different experimental setups. First, we demonstrate 5 $\times$  resolution enhancement on the setup in [25]. Next, we demonstrate the scalability of our technique by synthesizing multiplexed measurements on both static and dynamic biological data from [7] and achieve 4 $\times$  resolution improvement. In addition, the robustness of our technique to common experimental factors is quantified by evaluating the BNN-predicted uncertainties, including spatially varying aberrations, illumination misalignment, and phase wrapping artifacts. Mostly importantly, the results show that the selection of the training data indeed affects the confidence of the prediction, whose effect can be quantified by our UL framework. Specifically, we investigate the effect of limited training data due to spatial and temporal constraints and biological sample types. Furthermore, the BNN is shown to be reliable when trained and tested on different sample types and under different experimental configurations. The



**Fig. 1.** Overview of our reliable DL-based phase imaging technique. (a) Our technique opens up an expanded imaging attribute space, bypassing the conventional trade-off between FOV, resolution, and acquisition speed. (b) It uses five asymmetric illumination-coded intensities to encode large-SBP phase information. (c) A BNN is developed to make phase predictions and quantify the uncertainties of the model.

BNN-predicted uncertainties are shown to be indicative to the true error. Finally, a potential utility of our UL framework is explored in a time-series experiment to identify rare biological structures and phenomena.

## 2. METHOD

### A. Multiplexed Illumination for Large-SBP Phase Imaging

Our illumination multiplexing scheme combines the physical principles of DPC [18] and FPM [6] to encode high-resolution phase information across a wide FOV using a small number of intensity measurements. DPC is a phase microscopy technique that involves taking intensity measurements using asymmetric illumination [26]. Under the first Born approximation, a brightfield intensity measurement is linearly related to a sample's permittivity contrast by a weak phase transfer function [18]. The distribution of the transfer function affects the quality of the phase retrieval and can be tuned by adjusting the illumination pattern. Most importantly, the transfer function contains missing frequencies along the axis of asymmetry for a given illumination pattern [18]. As a result, illumination patterns containing at least two axes of asymmetry are commonly used to ensure complete Fourier coverage. Several studies on the choice of illumination patterns have been performed based on the linear model [18,27]. A CNN-based technique has also been developed to optimize the illumination patterns using a data-driven framework [17]. It should be noted that the validity of the DPC model relies on the presence of a strong reference wave as in the brightfield measurements—the model no longer holds for darkfield measurements. Accordingly, the maximum resolution achievable by DPC is limited to  $2\times$  the objective NA.

To further extend the resolution by more than  $2\times$ , our technique adapts the principle of FPM. In FPM, intensities are measured with asymmetric illumination in both brightfield and darkfield. Next, an iterative algorithm that simultaneously retrieves phase information and carries out the synthetic aperture is implemented. As a result, this method can increase the resolution up to the sum of the illumination and objective NAs [6]. A major advantage of FPM is its ability to achieve both a wide FOV and a high resolution, i.e., a large SBP. However, its imaging throughput is limited by the long acquisition time imposed by the large data requirement. Specifically, the original sequential FPM (sFPM) requires taking hundreds of images since it requires scanning through all the controllable illumination angles one by one [6] [Fig. 1(b)]. The acquisition time can be shortened by illumination multiplexing in mFPM. In [20], a random multiplexing scheme is shown to achieve up to  $8\times$  data reduction. A hybrid multiplexing scheme that combines DPC in the brightfield with random multiplexing in the darkfield is shown to provide improved robustness in solving the phase-retrieval problem of mFPM [7]. However, all these FPM schemes are fundamentally limited by the conventional trade-off, which results in an undesirable quadratic increase in data requirement as the resolution increases [7].

Here, we develop a DL-augmented illumination multiplexing scheme that uses only five asymmetric illumination [Fig. 1(b)]. First, we design two brightfield patterns based on the DPC model with in-total two axes of asymmetry (every  $90^\circ$ ) to provide complete Fourier coverage within the brightfield limit. Next, we

design three darkfield patterns with in-total three axes of asymmetry (every  $120^\circ$ ) to further extend the Fourier coverage set by the sum of the illumination and objective NAs, same as in the FPM. A notable feature of the proposed scheme is that extending the resolution simply requires modifying the illumination scheme to use a larger darkfield pattern, without the need for additional measurements. This means that the data requirement remains the same as the resolution increases—bypassing the limitation imposed by conventional techniques. By doing so, we improve the throughput of the data acquisition process by trading off computational complexity. Specifically, the multiplexed measurements cannot be robustly inverted by existing model-based mFPM algorithms due to the severe ill-posedness of the inverse problem. We show that our proposed BNN-based algorithm overcomes this issue owing to its nonlinear multilayer structure.

### B. Uncertainty Learning Framework

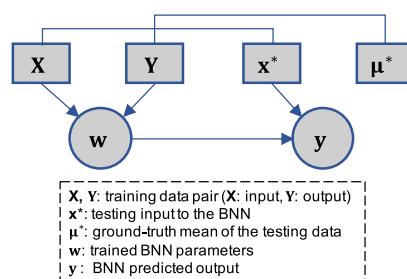
Our UL framework is built on the probabilistic view of neural networks [28]. The learned neural network differs from training to training, which in turn results in varied predictions. The variability stems from several stochastic processes involved in the training, such as random weight initialization [29], dropout [30], and stochastic-gradient-descent-type algorithms [31]. There are two ways to quantify the variabilities in a neural network, including the Bayesian [23] and frequentist [32] approaches. We outline both the approaches, provide the mathematical foundations for the Bayesian analysis, and then quantify uncertainties using both the Monte Carlo dropout [33] and the Deep Ensembles [32].

The BNN replaces the deterministic network weights with probability distributions over them [as illustrated in Fig. 1(a)]. To quantify the variability of a prediction  $\mathbf{y}$ , we model the predictive distribution  $p(\mathbf{y}|\mathbf{x}^*, \mathbf{X}, \mathbf{Y})$  given the test input  $\mathbf{x}^*$  through marginalization over all possible network weights  $\mathbf{w}$  that were learned from the training data  $(\mathbf{X}, \mathbf{Y}) = \{\mathbf{x}^t, \mathbf{y}^t\}_{t=1}^T$ :

$$p(\mathbf{y}|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}) = \int p(\mathbf{y}|\mathbf{x}^*, \mathbf{w})p(\mathbf{w}|\mathbf{X}, \mathbf{Y})d\mathbf{w}, \quad (1)$$

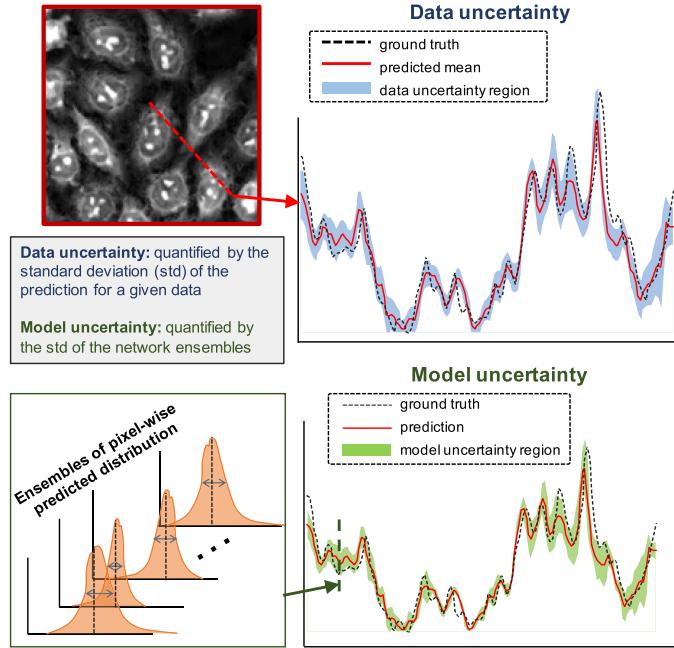
where Eq. (1) applies the conditional independence between the training and testing data, and can be visualized by the graphical model in Fig. 2. The posterior distribution  $p(\mathbf{w}|\mathbf{X}, \mathbf{Y})$  describes all possible *network weights* given the training data. The predictive distribution  $p(\mathbf{y}|\mathbf{x}^*, \mathbf{w})$  describes all possible *predictions* given the network weights  $\mathbf{w}$  and the testing input  $\mathbf{x}^*$  [Fig. 3(a), top]. By modeling  $p(\mathbf{w}|\mathbf{X}, \mathbf{Y})$  and  $p(\mathbf{y}|\mathbf{x}^*, \mathbf{w})$ , we can evaluate the model and data uncertainties, respectively.

To quantify the *data uncertainty*, we describe the probability distribution of the  $k$ th  $N$ -pixel random output of the BNN

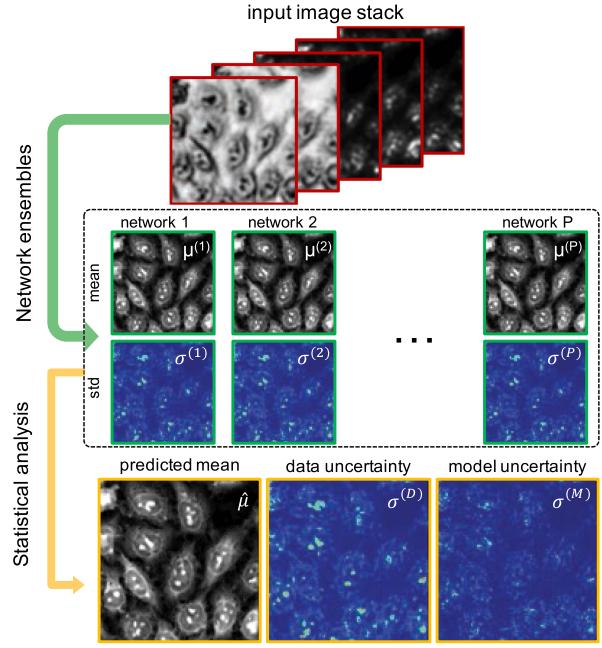


**Fig. 2.** Graphical model of our UL framework, which considers randomness in both the network weights  $\mathbf{w}$  and the predicted output  $\mathbf{y}$ .

## (a) Uncertainty quantification



## (b) Uncertainty visualization



**Fig. 3.** Overview of our UL framework. (a) The data uncertainty quantifies the effect of incomplete training data and is estimated via an uncertainty regularized loss function. The model uncertainty evaluates the stochasticity of neural network training and is estimated by network ensembles. (b) During testing, the direct output from the BNN consists of an ensemble of mean and standard deviation maps. Through statistical modeling, we obtain the final estimated phase, data, and model uncertainty maps.

(given the input  $\mathbf{x}^k$ ) by a multivariate Laplacian distributed likelihood function:

$$p(\mathbf{y}^k|\mathbf{x}^k, \mathbf{w}) = \prod_{i=1}^N p(y_i^k|\mathbf{x}^k, \mathbf{w}), \quad (2)$$

$$p(y_i^k|\mathbf{x}^k, \mathbf{w}) = \frac{1}{2\sigma_i^k} \exp\left(-\frac{|y_i^k - \mu_i^k|}{\sigma_i^k}\right), \quad (3)$$

where the output pixels (indexed by  $i$ ) are assumed to be independent, and  $\mu_i^k$  and  $\sigma_i^k$  denote the pixel-wise mean and standard deviation, respectively. It can be shown that the widely used mean absolute error (MAE) corresponds to this Laplacian model with a constant standard deviation assumed for the entire output [23]. By incorporating spatially *varying* standard deviations in our model, our BNN accounts for *inhomogeneous* noise and *shift-variant* model errors.

At the *training* stage, learning of the network weights is performed by minimizing the normalized negative log-likelihood function, i.e., the loss function  $L(\mathbf{w}|\mathbf{x}^t, \mathbf{y}^t)$ , given the training data  $(\mathbf{x}^t, \mathbf{y}^t)$ :

$$L(\mathbf{w}|\mathbf{x}^t, \mathbf{y}^t) = \frac{1}{N} \sum_{i=1}^N \left[ \frac{|y_i^t - \mu_i^t|}{\sigma_i^t} + \log(2\sigma_i^t) \right]. \quad (4)$$

$L(\mathbf{w}|\mathbf{x}^t, \mathbf{y}^t)$  consists of two parts: the first residual term resembles the MAE loss normalized by the pixel-wise standard deviation. The second is the *data uncertainty regularization* term. Most importantly, one does *not* need the ground-truth mean ( $\mu_i^t$ ) or the ground-truth standard deviation ( $\sigma_i^t$ ) for learning the uncertainty—minimizing  $L(\mathbf{w}|\mathbf{x}^t, \mathbf{y}^t)$  allows learning both using the sample pairs  $(\mathbf{X}, \mathbf{Y})$  taken from the random process. This is achieved by the structure of this loss function. Specifically, a large

residual error  $|y_i^t - \mu_i^t|$  will be regulated by a large standard deviation, which, in turn, increases the  $\log(2\sigma_i^t)$  term; the optimum can only be reached when the two terms are balanced. Training the BNN helps to not only find the optimal weights that explain all the data, but also quantify the individual *mismatch* between the data and the model as measured by the spread ( $\sigma_i^t$ ) in the network's output. At the predication stage, the BNN estimates both the mean and the standard deviation given the testing input, as illustrated in Fig. 1(c).

One approach to assess the *model uncertainty* is to use the dropout network [33]. Briefly, with dropout applied before every weight layer, a simple distribution  $q(\mathbf{w})$  is learned to provide a variational Bayesian approximation to the posterior  $p(\mathbf{w}|\mathbf{X}, \mathbf{Y})$ .

At the *prediction* stage, the model uncertainty is calculated by Monte Carlo dropout [33]. By using a Monte Carlo integration over  $P$  samples satisfying  $\mathbf{w}^{(p)} \sim q(\mathbf{w})$ , we can approximate the predictive distribution by a Laplacian mixture model:

$$p(\mathbf{y}|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}) \approx \int p(\mathbf{y}|\mathbf{x}^*, \mathbf{w}) q(\mathbf{w}) d\mathbf{w} \approx \frac{1}{P} \sum_{p=1}^P p(\mathbf{y}|\mathbf{x}^*, \mathbf{w}^{(p)}). \quad (5)$$

The variations in the distributions  $p(\mathbf{y}|\mathbf{x}^*, \mathbf{w}^{(p)})$  from the network ensembles are a consequence of the model uncertainty [Fig. 3(a), bottom].

The predicted mean  $\hat{\mu}_i$  of the  $i$ th pixel can be estimated by the unbiased minimum mean squared error estimator:

$$\hat{\mu}_i \equiv \mathbb{E}[y_i|\mathbf{x}^*, \mathbf{X}, \mathbf{Y}] \approx \frac{1}{P} \sum_{p=1}^P \mathbb{E}[y_i|\mathbf{x}^*, \mathbf{w}^{(p)}] \approx \frac{1}{P} \sum_{p=1}^P \mu_i^{(p)}, \quad (6)$$

where  $\mu_i^{(p)}$  is the predicted mean from the  $p$ th network, and  $\mathbb{E}$  denotes taking the expectation.

To provide a single, holistic measure of the uncertainty of the entire process, we quantify the overall uncertainty  $\hat{\sigma}_i$  by computing the pixel-wise variance (Var):

$$\begin{aligned}\hat{\sigma}_i^2 &\equiv \text{Var}(y_i | \mathbf{x}^*, \mathbf{w}) \\ &= \mathbb{E}[\text{Var}(y_i | \mathbf{x}^*, \mathbf{w})] + \text{Var}(\mathbb{E}[y_i | \mathbf{x}^*, \mathbf{X}, \mathbf{Y}]) \\ &\approx \frac{1}{P} \sum_{p=1}^P 2(\sigma_i^{(p)})^2 + \frac{1}{P} \sum_{p=1}^P (\mu_i^{(p)} - \hat{\mu}_i)^2 \\ &= (\sigma_i^{(D)})^2 + (\sigma_i^{(M)})^2,\end{aligned}\quad (7)$$

where the first equality follows the law of total variance, and the second one is derived from Eq. (5) and the Laplacian mixture model.  $\sigma_i^{(p)}$  denotes the pixel-wise standard deviation predicted from the  $p$ th network ensemble. Equation (7) shows that the overall data uncertainty  $\sigma_i^{(D)}$  is measured by the mean of the predicted variance; the model uncertainty  $\sigma_i^{(M)}$  is quantified by the variance of the predicted mean.

The second approach to quantify the uncertainties is the Deep Ensembles [32], in which multiple identical networks are trained under the same condition. A sufficient number of trained networks fully captures the variabilities of the model. We train eight networks to quantify the uncertainties. The model uncertainty is quantified by the same procedures in Eqs. (6) and (7).

Some examples of the predicted mean phase map, data uncertainty map, and model uncertainty map are shown in Fig. 3(b). Comparisons between the Monte Carlo dropout and the Deep Ensembles are provided in Supplement 1.

### C. BNN Structure

Our BNN follows the U-Net architecture owing to its versatility in solving image-to-image problems [34]. It takes the encoder-decoder structure with skip connections to preserve high-frequency features, as shown in Fig. 4. We made several modifications to perform uncertainty quantification. Mostly importantly, the output of the BNN contains two channels, including the predicted (mean) phase map and the data uncertainty standard deviation map. To achieve high resolution enhancement, we further adapt the generative adversarial network (GAN) [35]. We found that this GAN approach is needed to achieve 5 $\times$  resolution improvement of data from our setup. To achieve 4 $\times$  resolution improvement, however, we do *not* need

to use the GAN. The impact of GAN on the reliability of the prediction is analyzed in Section 3.C. Additional details about the network structure and training procedures are provided in Supplement 1. We have also made our implementation open-source, along with pre-trained weights and test sample data, on our GitHub project page [36].

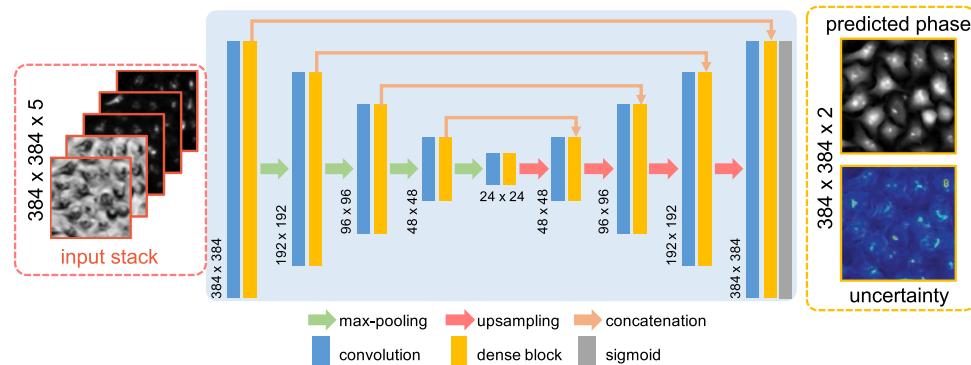
### D. Data Acquisition

Our technique is tested on two LED-array-based computational microscope setups, detailed in [7,25], and five different types of biological samples. First, we collect data on unstained HeLa cells prepared under two fixation conditions, including ethanol and formalin, on the setup in [25]. Depending on the fixation, unique morphologies can be observed in each sample, specifically, in the plasma membrane and nuclei regions. All images are captured with a 4 $\times$ , 0.1 NA objective (Nikon CFI Plan Achromat). Each data set consists of multiplexed data (two brightfield and three darkfield images) and the corresponding sFPM data (185 images). Both the multiplexed and sFPM data are captured with the same 0.41 illumination NA, providing a final resolution of 0.51 NA. Next, we validate our technique on data from [7]. The multiplexed measurements are synthesized by summing the single-LED images. We experimentally validate this procedure on the setup in [25] and find that the numerically synthesized multiplexed intensity closely matches the physically captured measurement since the LEDs are spatially and temporally incoherent. We test our method on both fixed U2OS, MCF10A, and dynamic live HeLa cell samples. Images were captured with a 4 $\times$ , 0.2 NA objective (CFI Plan Apo Lambda) at an illumination NA of either 0.5 or 0.6, which provide a final NA of 0.7–0.8. Each data set contains synthesized multiplexed and corresponding sFPM data. More details are provided in Supplement 1.

### E. Training and Test Data Configuration

We design three different training and testing data configurations in order to fully investigate the robustness of the BNN subject to different types of “limited data”, including unseen biological sample types, a limited FOV, and inaccessible temporal data.

In the first set of experiments, training data are taken from a single cell type; testing is then performed on several different cell types. In practice, different cells can produce out-of-distribution measurements that are not statistically “similar” to the training set. Differing from classification networks that are prone to



**Fig. 4.** BNN structure to perform UL. The main network takes the U-Net structure. The input takes five low-resolution multiplexed intensity images. The output predicts two-channel high-resolution phase and uncertainty maps.

testing errors from unseen object types, our network solves the inverse problem of an imaging model. As such, a properly trained network should be able to perform high-quality phase predictions and is robust to sample variations. We investigate how well the BNN can detect and quantify such abnormalities. In addition, we also study the network's robustness to variations in experimental setup.

In the second set of experiments, training data are taken from a limited FOV region, whereas testing data are taken from the entire FOV. This task is of practical importance because wide-field systems like FPM often suffer from spatially variant aberrations [37] and illumination misalignment [38]. These variations in the imaging path can change the intensity measurements significantly, such as contrast reversal, even when they are taken from the same sample, due to the interference effect. As a result, intensity measurements taken from FOV regions outside the training region can produce out-of-distribution data due to the limited training FOV. Differing from the model-based FPM approach, our data-driven BNN algorithm does not directly take any calibration information when constructing the network. Instead, the BNN needs to learn the spatially varying imaging model from the measurements and the ground truth phase. We will investigate the reliability of the BNN against these model variations.

In the final set of experiments, training data are taken from a limited observation time window from a time-series experiment. Dynamic biological processes can result in sample variations, which in turn affect the statistics of the intensity measurements, which may be inconsistent with the training set. We will assess the BNN's ability to make temporal predictions and quantify the uncertainty induced by the limited temporal data.

## F. Data Preprocessing

To obtain the ground truth phase for training, we first perform phase reconstruction using the sFPM algorithm [20]. To minimize model-mismatch-induced errors, we further perform algorithmic angle calibration using the algorithm in [38] and digitally correct for the aberrations using the algorithm in [20]. Additional preprocessing is performed to remove residual phase artifacts, including phase wrapping, a slowly varying background, and a large dynamic range. First, we perform phase unwrapping using the algorithm in [39]. Examples from this procedure are given in [Supplement 1](#). Next, the slowly varying background artifact is removed with a morphological-opening-based algorithm. Third, we perform phase dynamic range correction, which clips the 0.1% pixels having extreme values to be constant. Finally, the phase is linearly normalized to [0,1]. This processed phase map is then cropped into small patches for training. Still, the unwrapped phase contains residual isolated errors typically around large-phase or complex cellular features. This results in incorrect “phase labels” in the training data, which later affects the prediction. The impacts of incorrect labels and phase clipping on the uncertainties of the phase predictions are analyzed in detail in [Section 3.A](#).

To facilitate a later credibility analysis of the BNN output, we further quantify the noise present in the ground truth phase. Following [7], we measure the standard deviation in the background region and treat it as the intrinsic phase noise. We assume that the same noise level is uniformly distributed also across the sample (e.g., cell) regions. This noise level sets the tightest credible interval our BNN can provide; a detailed analysis is presented in [Section 3.C](#).

To preprocess the intensity measurements, background removal based on [20,40] is first performed, followed by dynamic range correction as in the ground truth phase preprocessing. Next, the full FOV is divided into small patches, which are resized with a cubic interpolation algorithm to match the size of the input image with the ground truth phase. For training, the matching phase and intensity patches are fed into the BNN. For testing, we apply an additional mean equalization to intensity patches taken from the *untrained* FOV region to alleviate the out-of-distribution effect. We find this procedure is essential to improve the BNN's generalization. Additional details about the preprocessing are provided in [Supplement 1](#).

## G. Data Analysis

We develop data analysis procedures to quantitatively relate the BNN predictions to Bayesian statistical reliability measures. Typical neural networks can only evaluate errors based on the ground truth, which is not possible for many practical problems. Here, we derive a set of metrics that do not require knowing the ground truth. Our analysis is based on the predictive Laplacian mixture model [Eq. (5)]. The probability density of the  $i$ th pixel to take the value  $y$  is

$$f_i(y) \equiv p(y_i = y | \mathbf{x}^*, \mathbf{X}, \mathbf{Y}) \approx \frac{1}{P} \sum_{p=1}^P \mathcal{L}(y; \mu_i^p, \sigma_i^p). \quad (8)$$

Accordingly, we define the *credible interval*  $A_i^\epsilon = [\mu_i - \epsilon, \mu_i + \epsilon]$  and its bound  $\epsilon$ . The corresponding *credibility*  $p_i^\epsilon$  is the predicted probability that the true mean  $\mu_i^*$  falls within  $A_i^\epsilon$ :

$$\begin{aligned} p_i^\epsilon &\equiv g_i(\epsilon) = \int_{\mu_i - \epsilon}^{\mu_i + \epsilon} f_i(y) dy \\ &= \frac{1}{P} \sum_{p=1}^P [F^p(\mu_i + \epsilon) - F^p(\mu_i - \epsilon)], \quad \text{for } y_i \in A_i^\epsilon, \end{aligned} \quad (9)$$

where  $F^p(\cdot)$  is the cumulative distribution function of the  $p$ th predicted Laplace distribution from the neural network ensembles. Another way to quantify the reliability is to calculate the bound  $\epsilon_i^p$  given a targeted credibility  $p$  and the predictive Laplacian mixture model; this can be computed by using the inverse function  $g_i^{-1}(\cdot)$ :

$$\epsilon_i^p = g_i^{-1}(p). \quad (10)$$

$g^{-1}(\cdot)$  does not have an elementary function that is approximated by the bisection method.

To ensure the predictive metrics in Eqs. (9) and (10) are indicative, we further characterize how well they are *calibrated* [41]. To quantify this, a standard procedure is to compute the *reliability diagram* that compares the *accuracy*, i.e., the empirical probability of the ground truth matching the predicted value, and the *credibility* [42]. Well-calibrated metrics should predict credibility similar to accuracy—the reliability diagram is diagonal. For a regression problem like ours, we adapt the modified reliability diagram [43], which compares the averaged credibility and the empirical accuracy. To generate a reliability diagram with  $M$  probability bins, we define the bin interval  $\Delta p = 1/M$  and the  $m$ th bin  $P_m$  bounded by  $p_{m-1}$  and  $p_m$ . The averaged credibility  $Cred(P_m, \epsilon)$  takes the mean over the set of pixels  $S_m^\epsilon$  having similar credibility within  $(p_{m-1}, p_m]$ :

$$\text{Cred}(P_m, \epsilon) = \frac{1}{|S_m^\epsilon|} \sum_{i \in S_m^\epsilon} p_i = \frac{1}{|S_m^\epsilon|} \sum_{p_i \in (p_{m-1}, p_m]} p_i, \quad (11)$$

where  $|S_m^\epsilon|$  measures the total number of pixels within the set. The empirical accuracy  $\text{Acc}(P_m, \epsilon)$  is defined as the fraction (empirical probability) of the pixels in set  $S_m^\epsilon$  in which the ground truth mean  $\mu_i^*$  is within the corresponding credible intervals  $A_i^\epsilon$ :

$$\text{Acc}(P_m, \epsilon) = \frac{1}{|S_m^\epsilon|} \sum_{i \in S_m^\epsilon} \mathbb{I}_{\{\mu_i^* \in A_i^\epsilon\}} = \frac{1}{|S_m^\epsilon|} \sum_{i \in S_m^\epsilon} \mathbb{I}_{\{\mu_i - \epsilon \leq \mu_i^* \leq \mu_i + \epsilon\}}. \quad (12)$$

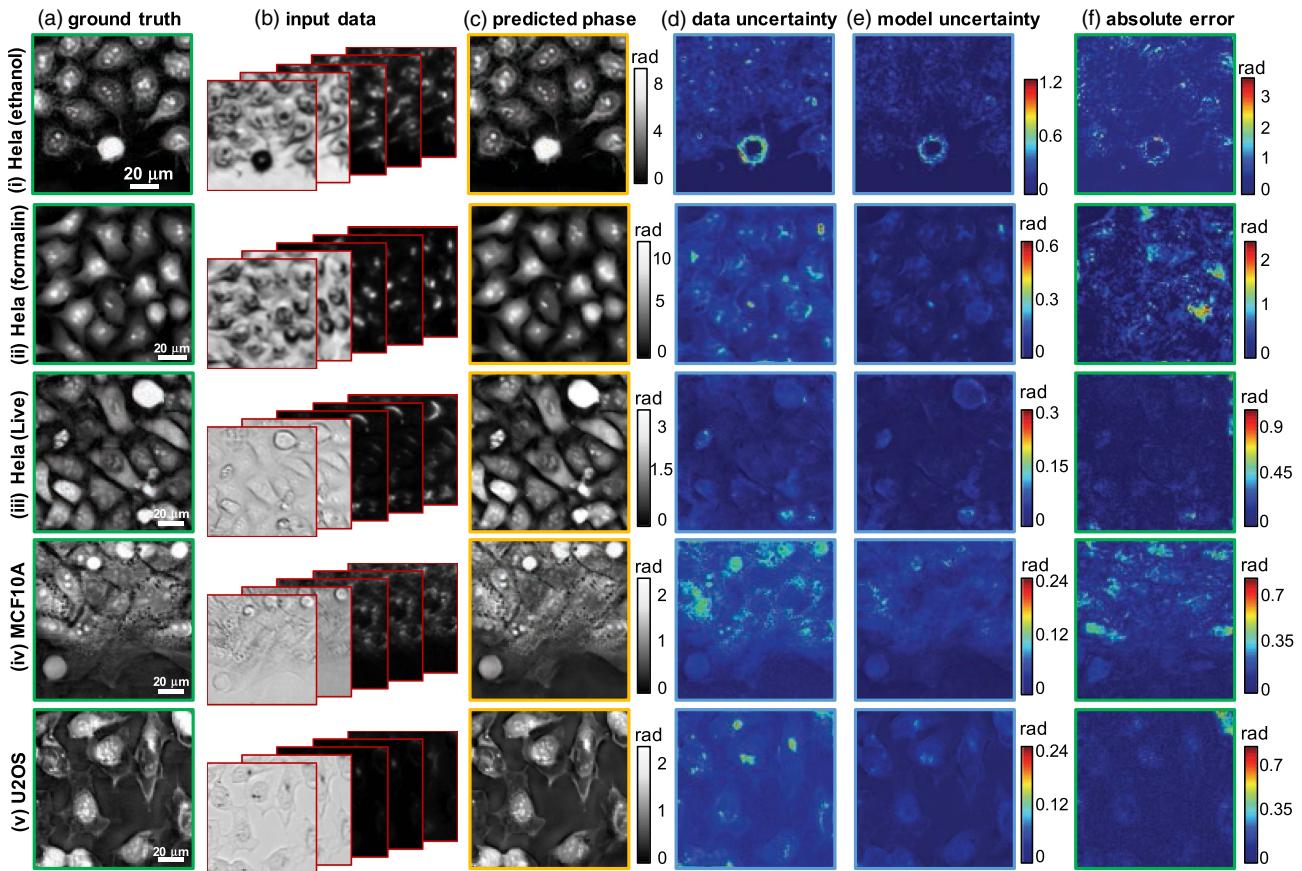
In practice, the ground truth mean is unknown and can only be approximated by the sFPM phase that is “noisy”, and so  $\text{Acc}(P_m, \epsilon)$  is influenced by the quality of the sFPM reconstruction. The bin interval sets the sampling interval in the reliability diagram and also affects the sample size in  $S_m^\epsilon$ . We use the minimum interval while ensuring sufficient sample size for reliable statistical calculation. Both the averaged credibility and empirical accuracy depend on the credible interval bound  $\epsilon$ . We assess our model using different  $\epsilon$  values in Section 3.C.

### 3. RESULTS

Our results are presented in the following order: first, we show that our technique provides high-resolution phase predictions and that the uncertainty maps are highly indicative to the true error. In addition, we show that the method is scalable to different sample types and is applicable to experimental setups with varying final resolution. Second, we present large-SBP phase prediction and show that the uncertainty maps allow quantifying the effect of out-of-distribution data due to a limited FOV. Third, we establish the reliability of our technique by performing statistical analysis. Finally, we demonstrate time-series predictions and show that UL can facilitate the discovery of spatially and temporally rare biological features and events.

#### A. Scalable Illumination-Coding-Based DL Phase Imaging

Our illumination coding scheme is highly scalable to large-SBP applications since it always uses five multiplexed measurements for achieving different resolutions. Experiments are performed on five cell types captured with two microscope setups and three different resolutions are achieved. Specifically, Figs. 5(i) and 5(ii) are obtained with the setup in [25] and achieve resolution



**Fig. 5.** High-resolution phase estimation from DL-augmented coded measurements. (a) The ground truth phase obtained from the sFPM. (b) The input to the neural network consists of five low-resolution intensity images, including two brightfield and three darkfield images. Our BNN prediction includes (c) phase, (d) data uncertainty, and (e) model uncertainty. (f) The absolute error is calculated between the predicted and ground truth phases. The uncertainty maps are highly correlated with the error maps, demonstrating the predictive power of our UL framework. Unlike existing FPM techniques, our method requires the same number of measurements when the final resolution increases. HeLa cells fixed in (i) ethanol and (ii) formalin are imaged with a  $4 \times 0.1$  NA objective and reconstructed at a resolution of 0.5 NA. (iii) Live HeLa and (iv) fixed MCF10A cells are imaged with a  $4 \times 0.2$  NA objective and reconstructed at a resolution of 0.8 NA. (v) Fixed U2OS cells are imaged with a  $4 \times 0.2$  NA objective and reconstructed at a resolution of 0.7 NA.

enhancement from 0.1 NA to 0.51 NA. Figures 5(iii)–5(v) are obtained with the setup in [7]; Figs. 5(iii) and 5(iv) enhance resolution from 0.2 NA to 0.8 NA, and Fig. 5(v) from 0.2 NA to 0.7 NA. First, we present results from training individual network for each cell type. Without any hyper-parameter tuning, the same network structure is applicable to different samples captured on different setups. Next, we show that the BNN trained with a single cell type is generalizable to other “unseen” cell types.

Example multiplexed intensity measurements are shown in Fig. 5(b). Our BNN is able to consistently provide high-quality phase predictions, as shown in Fig. 5(c). To evaluate a BNN-predicted phase, we first compare it with the phase from sFPM [Fig. 5(a)] and then compute the pixel-wise absolute error map [Fig. 5(f)]. Adding the additional uncertainty prediction in the BNN does not degrade the phase predictions as compared with the CNN approach (see *Supplement 1*). To demonstrate the need for using the DL method to overcome the ill-posedness of the phase-retrieval problem, we compare our results with those from two state-of-the-art model-based algorithms using the same multiplexed measurements. The linear DPC model [18] can only recover phase with limited resolution, whereas the mFPM algorithm [7] results in high-frequency artifacts in the recovered phase (see *Supplement 1*).

Next, we inspect the BNN-predicted data uncertainty [Fig. 5(d)] and model uncertainty maps [Fig. 5(e)]. The regions where the BNN *potentially* makes larger errors are marked with higher uncertainties. We observe that the uncertainty maps generally match well the corresponding absolute error map. In addition, the predicted uncertainty values are about 1/3 of the absolute error. This is because for a Laplace distribution, “ $3\sigma$ ” closely approximates the credible interval bound with 95% credibility. This demonstrates the utility of the uncertainty maps as a direct measure of the accuracy of the neural network predictions. Further quantitative reliability analyses are discussed in Section 3.C. In addition, we observe that the data uncertainty is the dominant term in our experiments, which suggests that the incompleteness in the training data is the main source of error in the prediction. Indeed, our training data are only taken from a small region of the FOV, as further discussed in Section 3.B. The low model uncertainty indicates that the predicted phase (i.e., pixel-wise mean) does not vary much across different neural network ensembles. This suggests that phase predictions based on the multiplexed measurements can be performed *consistently*—the stochastic training process does not lead to unstable inference results. Furthermore, the high uncertainty regions consistently correspond to cellular features with large phase values. We attribute this to two primary sources of error. First, the phase clipping inevitably introduces unwanted saturation artifacts in the ground truth phase. Second, although we correct for phase wrapping artifacts when generating the ground truth, residual errors still exist. Due to the presence of these inconsistencies in the training data associated with the large-phase features, the trained BNN tends to flag such “abnormal” regions in the uncertainty output.

Our BNN is trained to solve an inverse problem. As such, a properly trained network learns to invert the physical model, which is independent of the type of object used in the training. To justify this proposition, we compare the results from the BNN trained from the same cell type and from a different cell type in Fig. 6. In general, the BNN is able to make high-quality phase predictions and is robust to the selection of sample type.

Nevertheless, slight degradation is observed in the phase predicted from the network trained from a different cell type. This is because different cell types have distinct morphological features that can result in different intensity measurements. If the training data do not fully capture the statistical variations in the measurements, less accurate phase predictions would be produced when the network input contains “out-of-distribution” measurements. Most importantly, the uncertainty map from the BNN can automatically detect such abnormalities in the data. As highlighted in Fig. 6, the uncertainty map remains highly indicative to the true absolute error regardless of the cell type being used for training and testing. Additional results to demonstrate the robustness of our BNN to both sample and setup variations are provided in *Supplement 1*.

## B. Large-SBP Phase Prediction and Uncertainty Quantification

Next, we present large-SBP phase prediction across a wide FOV. Our BNN is trained on small image patches. We perform phase and uncertainty predictions patch-by-patch. The full-FOV predictions in Figs. 7(a)–7(c) are obtained by stitching the patches using the alpha blending algorithm.

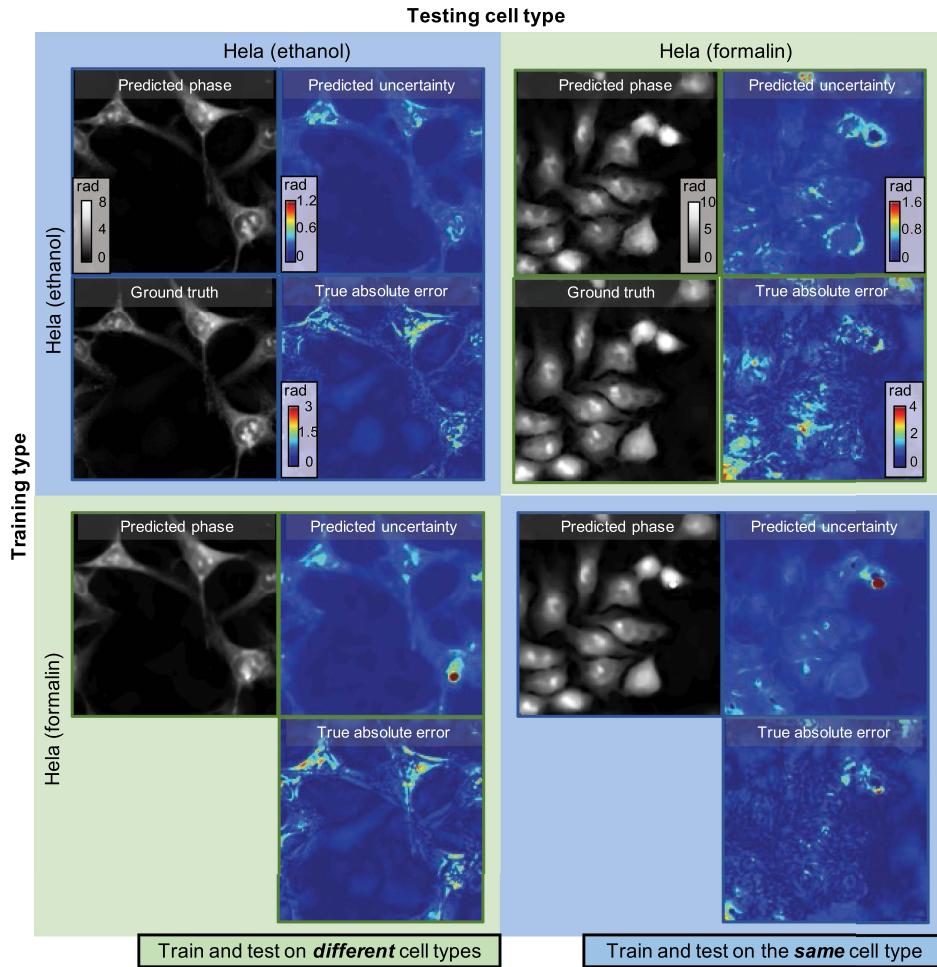
The full-FOV *model* uncertainty [Fig. 7(c)] allows critically assessing the robustness of our technique. We observe that the model uncertainty is low across the FOV except in small regions around the boundary. This verifies that the BNN can reliably make high-resolution phase predictions from the multiplexed measurements—the predicted mean does not vary much across different network ensembles. In the boundary regions, the measurements suffer from severe experimental errors that lead to higher variations in the predicted means.

The effect of out-of-distribution data due to the limited FOV is studied as follows: our *training* data are taken from a small central region ( $0.4 \text{ mm} \times 0.4 \text{ mm}$ ) from the full FOV of  $3.5 \text{ mm} \times 4.2 \text{ mm}$ , as shown in Fig. 7(d). In general, aberration degrades as the field angle increases (i.e., the distance away from the center). In addition, the LED illumination produces greater angle miscalibration [44] and background non-uniformity as the field angle increases. Both effects imply a greater degree of out-of-distribution as compared with the training data. Importantly, our UL approach allows predicting potential errors induced by the out-of-distribution data—the data uncertainty map predicts higher standard deviation at the peripheral FOV regions [Fig. 7(b)].

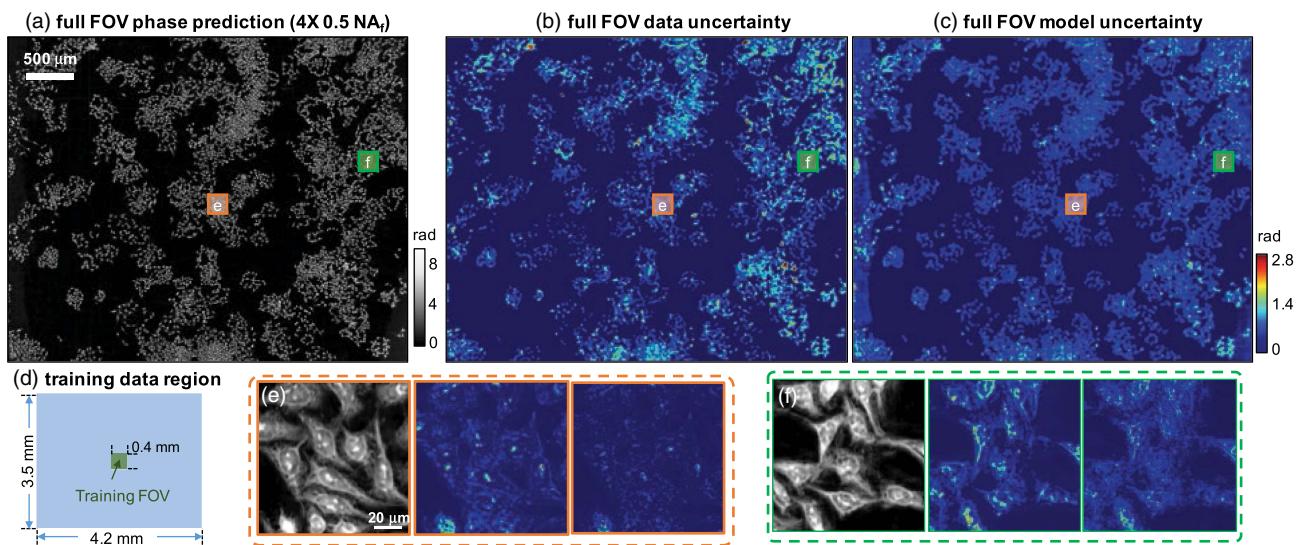
Identifying such data incompleteness *a posteriori* provides important feedback to improve the data pipeline in DL. Intuitively, introducing previously out-of-distribution data to the training can reduce the data uncertainty. In our case, more credible predictions can be made by training on more examples encompassing aberrations and angle miscalibration in other FOV regions, as verified by additional experiments detailed in *Supplement 1*.

## C. Quantitative Reliability Analysis

To provide a quantitative assessment to our prediction, we first calculate the *credibility map* from the predicted pixel-wise distribution. Given the bound  $\epsilon$  and the predicted mean  $\mu_i$  (at pixel  $i$ ), the credibility  $p_i^\epsilon$  [Eq. (9)] measures the BNN-predicted probability that the true mean falls in the credible interval  $A_i^\epsilon = [\mu_i - \epsilon, \mu_i + \epsilon]$ . To properly choose  $\epsilon$ , we consider the intrinsic noise in the phase reconstructed by sFPM by measuring



**Fig. 6.** BNN predictions under different training and testing data configurations. The network can robustly perform phase retrieval against variations in the sample type. The uncertainty map can reliably detect potential errors in the phase predictions and is consistent with the true error.



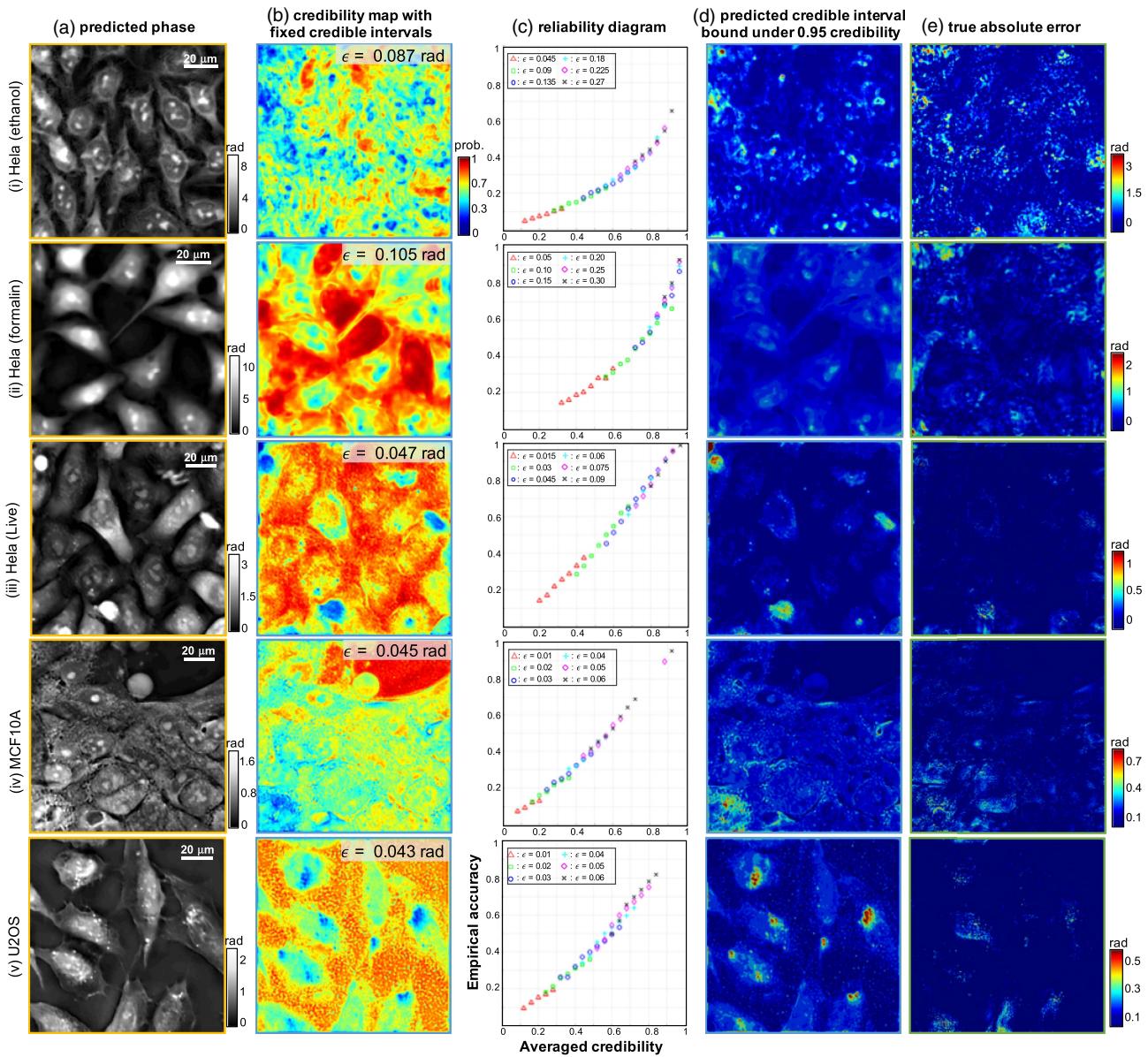
**Fig. 7.** Large-SBP phase prediction and uncertainty quantification. (a) Full-FOV phase prediction achieving 0.51 NA resolution across a  $4 \times$  FOV. (b) The data uncertainty map reliably identifies the out-of-distribution data corresponding to the peripheral FOV regions. (c) The model uncertainty is consistently low across the FOV except around the boundary, validating the robustness of our model. (d) The training data are taken only from the central  $0.4 \text{ mm} \times 0.4 \text{ mm}$  region. Panels (e) and (f) show the zoom-in of the predicted phase, data uncertainty, and model uncertainty of the region from the central FOV and the outer FOV, respectively.

the background standard deviation  $\sigma_{\text{background}}$ . We take this sFPM noise level as the credible interval bound ( $\epsilon = \sigma_{\text{background}}$ ) and compute the credibility pixel-by-pixel. The credibility map provides a direct quantification of how much one can trust the BNN-predicted phase. The credibility maps for the five samples and the credible interval bounds are shown in Fig. 8(b). As expected, less credible regions point to the “abnormal” regions where phase clipping or wrapping artifacts are likely present in the training data.

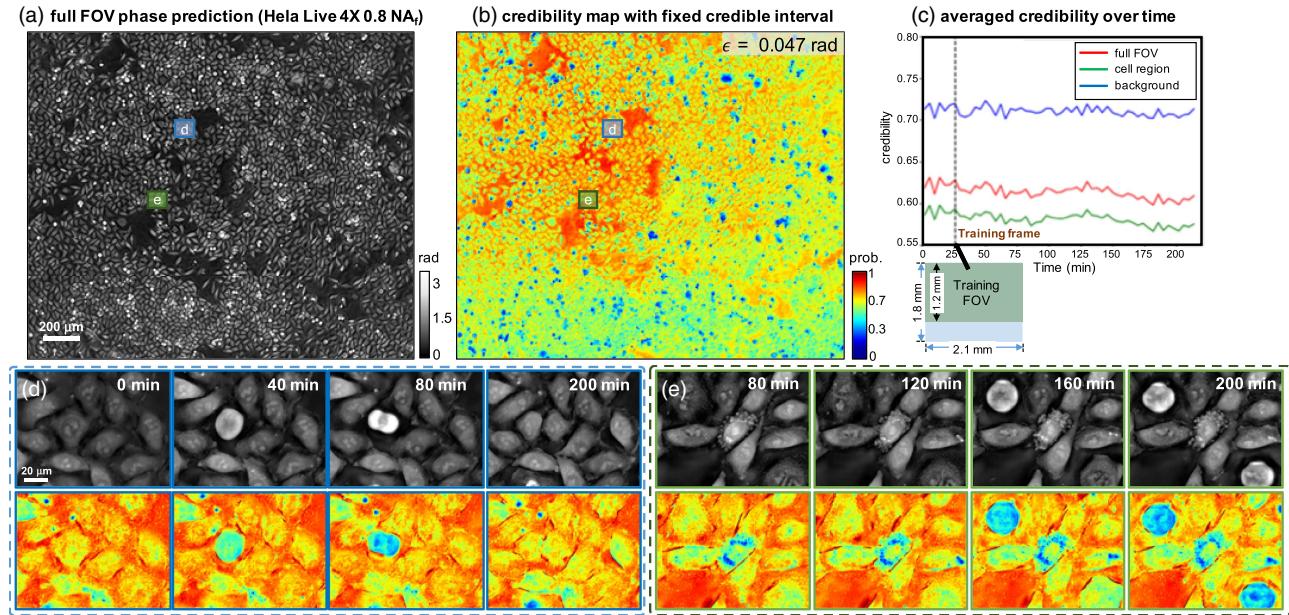
Alternatively, we evaluate the credible interval bound yielding a desired credibility. The bound  $\epsilon_i^p$  (at pixel  $i$ ) is computed using Eq. (10). By setting a constant  $p = 0.95$  (i.e., 95%) credibility across the whole image, we compute the predicted credible interval bound map as shown in Fig. 8(d). We observe that the credible interval bound map generally encompasses the corresponding

true absolute error map [Fig. 8(e)]. These results match well our previous observations on the predicted uncertainty maps.

Finally, we assess how well our UL framework is calibrated. We generate the reliability diagram [Fig. 8(c)] by computing the averaged credibility [Eq. (11)] and the approximated accuracy [Eq. (12)]. We set a probability bin interval of  $\Delta p = 0.04$  and use six credible interval bounds ( $\epsilon$ ). The first two cases [Figs. 8(i) and 8(ii)] with GAN included both show slightly overconfident predictions, as indicated by the curves below the diagonal. The other three cases [Figs. 8(iii)–8(v)] without GAN provide better calibrated predictions since the curves closely follow the diagonal. Besides the difference in the BNN structures, the first two cases have  $\sim 3\times$  stronger phase resulting in more phase-clipping-induced errors, and  $\sim 2\times$  higher intrinsic noise in the ground truth. Since the estimated empirical accuracy is also



**Fig. 8.** Reliability analysis of our predictions. (a) The predicted phase. (b) The credibility map calculated under the credible interval bound set by the intrinsic noise in sFPM. The less credible regions match with the out-of-distribution data containing phase clipping and wrapping artifacts. (c) The reliability diagram computed by comparing the averaged credibility and empirical accuracy show that (i–ii) are slightly overconfident whereas (iii–v) are well calibrated. (d) The predicted credible interval bound under 95% credibility correlates well with the corresponding true absolute error in (e).



**Fig. 9.** Time-series phase and credibility prediction. A representative frame from (a) the full-FOV phase prediction achieving 0.8 NA across 4 $\times$  FOV and (b) a credibility map with a credible interval bound of  $\epsilon = 0.047$  rad. (c) The training data are taken from the upper 3/4 of the FOV at the 26 min frame. The averaged credibility are calculated over time on the whole FOV (red), the cell region (green), and the background (blue). This allows quantifying the “temporal decorrelation” induced by the temporal dynamics. (d, e) Spatially and/or temporally rare events, including cell mitosis and apoptosis, which result in out-of-distribution data during prediction, are automatically discovered by our BNN. The full time-series prediction is provided in the movie in [Visualization 1](#).

influenced by the quality of the ground truth, the lower quality ground truth phase in the first two cases could also contribute to the less calibrated predictions. Methods to improve the calibration of the BNN is an active area of research [41] and will be developed in our future work.

#### D. Time-Series Large-SBP Phase and Credibility Prediction

Our technique is also applicable to imaging dynamic samples. Figure 9 shows time-series predictions made by training the BNN using data only from a single time frame. We train the BNN using the upper 3/4 of the FOV at the 26 min frame and perform full-FOV predictions on the rest of the time frames. An example FOV phase prediction is shown in Fig. 9(a). The reliability of the temporal predictions is further quantified by calculating the credibility maps over time. An example credibility map is shown in Fig. 9(b). As expected, the BNN is credible across the entire *trained* FOV region and less credible over the *untrained* region, matching our previous observations.

To quantify the reliability over time, we calculate the averaged credibility over the full FOV, the cell, and the background regions [Fig. 9(c)]. The averaged credibility fluctuates within a small range. The credibility for the cell regions slowly decays over time; this can be explained by the reason that the temporal dynamics gradually induce more “dissimilar” out-of-distribution data. Our BNN enables quantifying such “temporal decorrelation”.

Next, we zoom in on two small regions where cell divisions undergo over time [Figs. 9(d) and 9(e)]. In both cases, credibility drops when the cells present significant morphological changes during mitosis, and increases back to the “normal” level immediately after the process is over. More examples are shown in the

movie in [Visualization 1](#). As the cells become more globular during mitosis, the phase values grow significantly and often result in phase wrapping errors in the training phase data. In Fig. 9(e), a cell undergoes apoptosis and presents distinct morphological structures. Similar to our previous observations, the BNN consistently identifies these spatially and temporally rare features by “flagging” them as being less credible.

## 4. CONCLUSION

We have presented a physics-guided DL framework for large-SBP phase imaging. Our technique enables high-resolution phase inference across a wide FOV using only five asymmetric illumination-coded intensity measurements. Our results show that this BNN-based technique can effectively learn the underlying physical model. Once trained, the BNN can robustly solve the phase-retrieval problem and is generalizable to different samples. Further, we have developed an uncertainty quantification framework that allows critically assessing the reliability of the BNN predictions. Specifically, we have applied our UL approach to evaluate the robustness of our illumination coding and DL phase estimation model. In addition, we have also quantified the effect of common experimental errors using the predicted uncertainties. Furthermore, we have shown that applying the UL enables discovering the incompleteness in the training data and quantifying the associated out-of-distribution testing errors. Finally, the predicted credibility map has been shown to be useful for identifying spatially and temporally rare biological phenomena, and for characterizing the “temporal decorrelation” in dynamic processes. We believe this UL framework is widely applicable to many emerging DL-based scientific and biomedical imaging applications where critical assessment to the DL inference is essential.

**Funding.** National Science Foundation (NSF) (1813848); National Institutes of Health (NIH) (R21GM128020).

**Acknowledgment.** We thank Dr. Ji Yi for providing the samples in our experiments, and Joseph Greene and Alex Matlock for helpful discussions.

## REFERENCES

1. A. W. Lohmann, R. G. Dorsch, D. Mendlovic, Z. Zalevsky, and C. Ferreira, "Space-bandwidth product of optical signals and systems," *J. Opt. Soc. Am. A* **13**, 470–473 (1996).
2. A. W. Lohmann, "Scaling laws for lens systems," *Appl. Opt.* **28**, 4996–4998 (1989).
3. W. Lukosz, "Optical systems with resolving powers exceeding the classical limit," *J. Opt. Soc. Am.* **56**, 1463–1471 (1966).
4. W. Lukosz, "Optical systems with resolving powers exceeding the classical limit ii," *J. Opt. Soc. Am.* **57**, 932–941 (1967).
5. K. Wicker and R. Heintzmann, "Resolving a misconception about structured illumination," *Nat. Photonics* **8**, 342–344 (2014).
6. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nat. Photonics* **7**, 739–745 (2013).
7. L. Tian, Z. Liu, L.-H. Yeh, M. Chen, J. Zhong, and L. Waller, "Computational illumination for high-speed in vitro Fourier ptychographic microscopy," *Optica* **2**, 904–911 (2015).
8. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," *Optica* **4**, 1117–1125 (2017).
9. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.* **7**, 17141–17149 (2018).
10. T. Nguyen, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, "Deep learning approach for Fourier ptychography microscopy," *Opt. Express* **26**, 26470–26484 (2018).
11. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica* **5**, 803–813 (2018).
12. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica* **5**, 1181–1190 (2018).
13. Y. Wu, Y. Rivenson, Y. Zhang, Z. Wei, H. Günaydin, X. Lin, and A. Ozcan, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," *Optica* **5**, 704–710 (2018).
14. R. Horstmeyer, R. Y. Chen, B. Kappes, and B. Judkewitz, "Convolutional neural networks that teach microscopes how to image," arXiv:1709.07223 (2017).
15. B. Diederich, R. Wartmann, H. Schadwinkel, and R. Heintzmann, "Using machine-learning to optimize phase contrast in a low-cost cellphone microscope," *PLoS One* **13**, e0192937 (2018).
16. A. Robey and V. Ganapati, "Optimal physical preprocessing for example-based super-resolution," *Opt. Express* **26**, 31333–31350 (2018).
17. M. Kellman, E. Bostan, N. Repina, and L. Waller, "Physics-based learned design: optimized coded-illumination for quantitative phase imaging," *IEEE Transactions on Computational Imaging (Early Access)* (2019), <https://doi.org/10.1109/TCI.2019.2905434>.
18. L. Tian and L. Waller, "Quantitative differential phase contrast imaging in an LED array microscope," *Opt. Express* **23**, 11394–11403 (2015).
19. T. R. Hillman, T. Gutzler, S. A. Alexandrov, and D. D. Sampson, "High-resolution, wide-field object reconstruction with synthetic aperture Fourier holographic optical microscopy," *Opt. Express* **17**, 7873–7892 (2009).
20. L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for Fourier ptychography with an LED array microscope," *Biomed. Opt. Express* **5**, 2376–2389 (2014).
21. E. Bostan, M. Soltanolkotabi, D. Ren, and L. Waller, "Accelerated Wirtinger flow for multiplexed Fourier ptychographic microscopy," arXiv:1803.03714 (2018).
22. P. Chen and A. Fannjiang, "Coded aperture ptychography: uniqueness and reconstruction," *Inverse Probl.* **34**, 025003 (2018).
23. A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Advances in Neural Information Processing Systems* (2017), pp. 5580–5590.
24. A. D. Kiureghian and O. Ditlevsen, "Aleatory or epistemic? Does it matter?" *Struct. Saf.* **31**, 105–112 (2009).
25. R. Ling, W. Tahir, H.-Y. Lin, H. Lee, and L. Tian, "High-throughput intensity diffraction tomography with a computational microscope," *Biomed. Opt. Express* **9**, 2130–2141 (2018).
26. S. Mehta and C. Sheppard, "Quantitative phase-gradient imaging at high resolution with asymmetric illumination-based differential phase contrast," *Opt. Lett.* **34**, 1924–1926 (2009).
27. Y. Fan, J. Sun, Q. Chen, X. Pan, L. Tian, and C. Zuo, "Optimal illumination scheme for isotropic quantitative differential phase contrast microscopy," arXiv:1903.10718 (2019).
28. Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature* **521**, 452–459 (2015).
29. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International Conference on Artificial Intelligence and Statistics* (2010), Vol. 9, pp. 249–256.
30. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
31. L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'10* (2010), pp. 177–186.
32. B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems* (2017), pp. 6402–6413.
33. Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: representing model uncertainty in deep learning," in *International Conference on Machine Learning* (2016), pp. 1050–1059.
34. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.
35. P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," CoRR abs/1611.07004 (2016).
36. Y. Xue, S. Cheng, Y. Li, and L. Tian, <https://github.com/bu-cisl/illumination-coding-meets-uncertainty-learning>.
37. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," *Opt. Express* **22**, 4960–4972 (2014).
38. R. Eckert, Z. F. Phillips, and L. Waller, "Efficient illumination angle self-calibration in Fourier ptychography," *Appl. Opt.* **57**, 5434–5442 (2018).
39. D. C. Ghiglia and L. A. Romero, "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," *J. Opt. Soc. Am. A* **11**, 107–117 (1994).
40. L.-H. 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).
41. V. Kuleshov, N. Fenner, and S. Ermon, "Accurate uncertainties for deep learning using calibrated regression," arXiv:1807.00263 (2018).
42. A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning* (2005).
43. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, and M. Rocha-Martins, "Content-aware image restoration: pushing the limits of fluorescence microscopy," *Nat. Methods* **15**, 1090–1097 (2018).
44. J. Sun, Q. Chen, Y. Zhang, and C. Zuo, "Efficient positional misalignment correction method for Fourier ptychographic microscopy," *Biomed. Opt. Express* **7**, 1336–1350 (2016).