



HAL
open science

A physics-inspired deep learning framework for an efficient fourier ptychographic microscopy reconstruction under low overlap conditions

Lyes Bouchama, Bernadette Dorizzi, Yaneck Gottesman, Jacques Klossa

► To cite this version:

Lyes Bouchama, Bernadette Dorizzi, Yaneck Gottesman, Jacques Klossa. A physics-inspired deep learning framework for an efficient fourier ptychographic microscopy reconstruction under low overlap conditions. *Sensors*, 2023, 23 (15), pp.6829. 10.3390/s23156829 . hal-04397977

HAL Id: hal-04397977

<https://hal.science/hal-04397977>

Submitted on 16 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

A Physics-Inspired Deep Learning Framework for an Efficient Fourier Ptychographic Microscopy Reconstruction under Low Overlap Conditions

Lyes Bouchama ^{1,2,*}, Bernadette Dorizzi ¹ , Jacques Klossa ² and Yaneck Gottesman ¹ 

¹ Samovar, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France; bernadette.dorizzi@telecom-sudparis.eu (B.D.); yaneck.gottesman@telecom-sudparis.eu (Y.G.)

² TRIBVN/T-Life, 92800 Puteaux, France; jklossa@tribvn.com

* Correspondence: lbouchama@tribvn.com

Abstract: Two-dimensional observation of biological samples at hundreds of nanometers resolution or even below is of high interest for many sensitive medical applications. Recent advances have been obtained over the last ten years with computational imaging. Among them, Fourier Ptychographic Microscopy is of particular interest because of its important super-resolution factor. In complement to traditional intensity images, phase images are also produced. A large set of \mathcal{N} raw images (with typically $\mathcal{N} = 225$) is, however, required because of the reconstruction process that is involved. In this paper, we address the problem of FPM image reconstruction using a few raw images only (here, $\mathcal{N} = 37$) as is highly desirable to increase microscope throughput. In contrast to previous approaches, we develop an algorithmic approach based on a physics-informed optimization deep neural network and statistical reconstruction learning. We demonstrate its efficiency with the help of simulations. The forward microscope image formation model is explicitly introduced in the deep neural network model to optimize its weights starting from an initialization that is based on statistical learning. The simulation results that are presented demonstrate the conceptual benefits of the approach. We show that high-quality images are effectively reconstructed without any appreciable resolution degradation. The learning step is also shown to be mandatory.

Keywords: Fourier ptychographic microscopy; deep learning; deep image prior



Citation: Bouchama, L.; Dorizzi, B.; Klossa, J.; Gottesman, Y. A Physics-Inspired Deep Learning Framework for an Efficient Fourier Ptychographic Microscopy Reconstruction under Low Overlap Conditions. *Sensors* **2023**, *23*, 6829. <https://doi.org/10.3390/s23156829>

Academic Editor: Chau-Jern Cheng

Received: 27 June 2023

Revised: 20 July 2023

Accepted: 28 July 2023

Published: 31 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Two-dimensional observation of biological samples at hundreds of nanometers resolution or even below is of high interest for many research activities such as fundamental biology, drug discovery, or other medical applications. This has motivated numerous studies in optical microscopy to push its actual limits regarding the delivered images' ultimate resolution caused by light diffraction. Additional access to information such as sample optical thickness ($2D^{1/2}$) or even a complete 3D representation of the sample is also highly desired.

In such context, different computational microscopy approaches have made important progress over the last decade [1,2]. For these approaches, the image of the sample is obtained after solving an inverse problem. Most of available methods exploit multiple measurements of the sample under different experimental conditions [3–6]. Acquired data (or raw images) are then exploited computationally to reconstruct the information that is not directly accessible. Among available techniques, Fourier Ptychographic Microscopy (FPM) is of particular interest [7,8]. It is based on a conventional optical microscope with few modifications in the employed illumination source. An impressive super-resolution factor up to ~ 5 is accessible (demonstrated experimentally) using 225 raw images together with a quantitative assessment of absorption and local optical thickness of samples. Hence, such microscopy is a strong candidate for different applications. In addition, the detection

of rare events is facilitated by the intrinsic increase in the microscope's space-bandwidth product (SBP).

Since the FPM acquisition time is proportional to the number of raw images that are captured, algorithmic approaches enabling good image reconstructions with few images are highly desirable. The first proposed method, from Waller's group [9,10] consisted in multiplex coding illumination. In this approach, the specimen is simultaneously probed with different angular plane waves. The different regions in Fourier domain that are captured by the camera in a single shot are then separated with a sort of source separation algorithm embedded in the reconstruction algorithm. Such method has two advantages: the number of images that needs to be recorded for a complete coverage of full targeted spectral region is much reduced; the total optical power used to probe the sample is increased in proportion with the number of LEDs used simultaneously, leading to a possible reduction in the camera exposure time. However, the exact consequences of such coding on reconstructed image quality is not yet clear since the information acquired with a given camera dynamic needs to be distributed over large spectral regions. Alternative approaches [11–14] have also been investigated where the reconstruction process is obtained from training a deep neural network (DNN). Here, the training process aims at approximating a general underlying relation between the raw captured images and reconstructed ones. Reconstruction in a deep-learning framework has been mainly motivated by many additional foreseen advantages such as faster reconstructions brought by GPUs in feed-forward calculations (inference), or better noise tolerances. Successful DNN reconstructions have already been obtained when a large number of raw images are used. However the previous published results that have been obtained experimentally or with simulations are not yet completely satisfactory when only a few raw images are exploited. For example, artifacts often appear in the high frequencies of reconstructed images although perceptually close to images reconstructed with the original FPM method.

The situation is about to change with the advent of deep image prior (DIP) networks [15,16]. In such a scheme, a neural network is used as a prior to solve inverse problems. One of the remarkable specificities of DIP is that it does not exploit any prior training data. Image characteristics are captured by the structure of a convolutional image generator rather than by any previously learned capabilities. In practice, the neural architecture introduces some regularization into the classic inversion optimization scheme. DIP has been proved to be successful in many imaging application fields such as noise reduction, super-resolution, and inpainting. More recently, image reconstruction with improved axial resolution has been obtained in the field of optical diffraction tomography [17,18]. However, and in contrast to the training approach, DIP approaches lack generalization capabilities since the solution is specific to each image considered.

In this paper, we address the difficult problem of FPM reconstruction with few raw images in a deep-learning context. The method that we propose will go beyond the two previous trends; more precisely, we here introduce a general method based on DIP formalism. In the first step, the DIP network is trained for reconstruction. In the second step, its weights are further optimized in order to obtain a precise estimation of the sought solution that is specifically attached to a given set of raw images. The DIP network is fully physics-informed across the two steps in the sense that the exact forward model of FPM image formation and light-sample interaction is introduced in its loss function [19]. The principle of such a reconstruction scheme is demonstrated with the help of simulations. Reconstruction with a super-resolution factor ~ 5 with no appreciable artifact in the high frequencies is achieved with only 37 raw images. A particular microscope illumination LED matrix pattern is considered in the simulations. Its spatial arrangement has been chosen in order to pave the Fourier space with as low a number of LEDs as possible. The approach is general and applicable to other illumination patterns such as the classic rectangular LED board. It can also open interesting perspectives for FPM reconstruction under multiplex-coded illumination to reduce further the number of captured images that are required.

This paper is organized as follows. The basic principle of FPM is briefly recalled in Section 2. Emphasize is put on the reconstruction problem when only a few low-resolution images are acquired. It is the key problem that is addressed and solved in Section 3 with the proposed cDIP-LO architecture. To this end, a unified DNN model is introduced, allowing both learning and optimization steps. In Section 4, the performance of the complete model is evaluated and discussed. With the help of simulation results, we demonstrate that both low and high frequencies are correctly reconstructed. The FPM configuration that is considered is associated with a super-resolution factor ~ 5 . The number of LR images exploited is reduced from 130 (overlap $\sim 60\%$) to 37 (overlap $\sim 10\%$). Section 5 is devoted to the conclusion and perspectives of the work.

2. Fourier Ptychographic Microscopy Principle

FPM relies on an optical microscope setup [7,8,20] in which the traditional illumination source is replaced by an LED matrix array (see Figure 1a). Each LED of the matrix is assimilated to a quasi-monochromatic punctual source [21]. Its distance to the sample is far enough to approximate the incident light with a plane wave whose wave-vector \mathbf{k} varies with its spatial position. Changing the i th LED that is turned on permits us to probe the sample with varied \mathbf{k}^i .

Let us denote $U_{in}^i(x, y)$ and $U_{out}^i(x, y)$, respectively, as the electric field that is incident on the sample and the electric field exiting from the sample for the i th. Under thin sample approximation, the light–matter interaction is modeled with a complex mask T . The relation between $U_{in}^i(x, y)$ and $U_{out}^i(x, y)$ becomes

$$U_{out}^i(x, y) = U_{in}^i(x, y) \cdot T(x, y) \quad (1)$$

with $U_{in}^i(x, y) = Ae^{j(k_x^i \cdot x + k_y^i \cdot y)}$. The image formation realized by the microscope at the camera plane consists in a low pass filtering of $U_{out}^i(x, y)$. The electric field $U_{cam}^i(x, y)$ at the camera plane becomes

$$U_{cam}^i(x, y) = U_{out}^i\left(\frac{x}{g}, \frac{y}{g}\right) * C(x, y) \quad (2)$$

where g represents the magnification factor of the objective lens, C the point spread function and $*$ the convolution product. In Fourier domain,

$$\widehat{U}_{cam}^i(k_x, k_y) = \widehat{T}(g(k_x - k_x^i), g(k_y - k_y^i)) \cdot \widehat{C}(k_x, k_y) \quad (3)$$

where \widehat{T} and \widehat{C} represents the Fourier transform of T and C . For an objective lens that is aberration free, $\widehat{C}(k_x, k_y) = 1$ in a region delimited by a disk of radius $r = \frac{2\pi}{\lambda} \cdot NA$ centered at $(0, 0)$, and 0 outside, where NA represents the numerical aperture of the objective lens. Equation 3 defines the forward model that describes the image formation of the sample illuminated with the i th LED. It is interesting to observe that the i th LED enables the grabbing of the region in the spectral domain that is delimited by a disk of radius r and centered on (k_x^i, k_y^i) . Figure 1b illustrates the position of these regions for the LED matrix that will be considered for simulations. It is composed of 37 LEDs arranged with a circular pattern.

Provided the phase is preserved, the many U_{cam}^i fields can be readily assembled in the Fourier domain in order to recover the sample spectrum over a region that is much larger than the one directly accessible to the objective lens. This constitutes the heart of the synthetic aperture mechanism exploited by FPM. This permits us to recover the sample with an enhanced resolution. The super-resolution factor γ that is attained is theoretically only imposed by the LED matrix layout, and hence the spectral region that is paved.

The i th image captured by the camera is

$$I^i(x, y) = |U_{cam}^i(x, y)|^2 \quad (4)$$

As a result, the phase is inevitably lost during the capture of the different LR images. It is hence mandatory to reconstruct the missing phase information. Different iterative approaches are available such as the Gershberg–Saxton's [22,23], ePIE [24] or EPRY [25] algorithms. Let us denote $\widehat{\mathcal{T}}$ as an approximated solution and I^i as the low-resolution image deduced after applying the direct model on $\widehat{\mathcal{T}}$ as indicated below:

$$\sqrt{I^i(x, y)} = |F^{-1}(\widehat{\mathcal{T}}(k_x, k_y) \cdot \widehat{C}(k_x - k_x^i, k_y - k_y^i))| \quad (5)$$

with F^{-1} being the Fourier inverse transform. The searched solution is hence found by minimizing the error function.

$$\mathcal{L} = \sum_i \left| \sqrt{I^i(x, y)} - \sqrt{I^i(x, y)} \right|^2 \quad (6)$$

With I^i being the different measured images. The above-mentioned algorithms proceed iteratively and $\widehat{\mathcal{T}}$ is reconstructed progressively with descent-gradient calculations. These calculations operate by exploiting images I^i , sweeping LEDs i from 1 to \mathcal{N} . The solution $\widehat{\mathcal{T}}$ is found after convergence. Super-resolved intensity and phase images are then calculated as the modulus and phase of the inverse Fourier transform of $\widehat{\mathcal{T}}$. In addition to the super-resolution and phase reconstruction abilities of these algorithms, the forward model can also integrate a digital wavefront correction mechanism (see [7] for further details). This allows an efficient digital focus correction of objects that are out of focus, as is highly desirable for many practical situations.

It is important to observe that the inverse problem to solve is ill-posed. This imposes some experimental conditions on the setup: The thermal noise needs to be sufficiently low as can be achieved by fixing adequate LED power. Furthermore, the number of LEDs (or LR images exploited) should be sufficient. Moreover, and since the descent gradient is performed iteratively, only partial experimental data are exploited (image I^i) at each step. The overlap factor Γ between two successive regions updated in Fourier Domain at iteration i and $i + 1$ must be above 50–60%. This question has been studied by many authors and $\Gamma > 60\%$ has been determined from simulations that were further confirmed experimentally [26,27]. These overlapping regions are indicated with the greyed regions in Figure 1. As a result, the Fourier domain needs to be paved with high redundancy, i.e., with many different images I^i . This limits the throughput of FPM. It is clear that the minimal theoretical quantity of data that should be acquired depends quadratically on the super-resolution factor γ . This γ factor is defined in the Fourier domain by the ratio between the surface to be reconstructed and the surface covered by the numerical aperture of the microscope lens. For example, and for $\gamma = 6$, only 36 different illumination angles (equivalently images) are theoretically necessary (for $\gamma \sim 0\%$). However, in order to ensure a correct convergence of the reconstruction algorithm, Γ should be above 60%. Sample data thus need to be acquired with high redundancy, leading to a penalizing throughput of the microscope. In [7], the experimental configuration used ensures $\gamma = 6$ and 225 illuminations were needed. As a result, the minimal time for the acquisition of one FOV in RGB colors is $\sim 1'30$ s considering only 20 ms of exposure time for each stack acquisition (in bright and even dark field conditions).

In the following, we consider an LED circular matrix \mathcal{M} (see Figure 1a) since its geometry permits to achieve a low Γ factor. This matrix is split into two distinct groups of LEDs denoted \mathcal{M}_A and \mathcal{M}_B with $\mathcal{M} = \mathcal{M}_A \cup \mathcal{M}_B$. Their layouts are indicated in Fourier domain Figure 1b and Figure 1c, respectively. The LED matrix \mathcal{M}_A permits us to pave a wide spectral region and hence to attain an important super-resolution factor. \mathcal{M}_A is composed of 37 LEDs and leads to an overlap factor $\Gamma \sim 10\%$. \mathcal{M}_A geometry has been designed to achieve the lowest overlap factor without any missing information in the Fourier domain. \mathcal{M}_B is composed of 93 complementary LEDs. When combined with the LEDs of \mathcal{M}_A , they permit us to attain $\Gamma = 60\%$ (refer to Figure 1's caption for further

details on the microscope configuration). As already indicated, the use of \mathcal{M}_A alone cannot permit us to achieve exploitable FPM HR image reconstruction using ePIE or related reconstruction algorithms because of Γ . This property is most likely general. Moreover, we observed (simulations not presented here) that the situation is unchanged even if a global gradient-descent algorithm [28] is used. That means that novel algorithmic reconstruction strategies are needed. We have therefore developed an approach relying on DNNs in order to overcome current limitations.

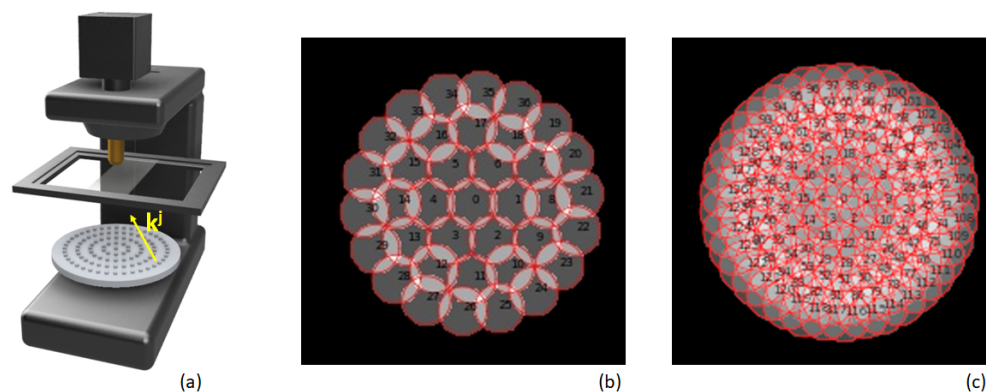


Figure 1. FPM reconstruction in Fourier domain. Schematic presentations of the microscope fitted with a LED matrix $\mathcal{M} = \mathcal{M}_A \cup \mathcal{M}_B$ (a), spectral region covered in Fourier Domain with \mathcal{M}_A (b) and with \mathcal{M}_B (c).

3. Physics-Informed cDIP-DNN Reconstruction Scheme

Our approach relies on a DIP implementation of the reconstruction, seen as an inversion of the forward FPM model described in Section 2.

DIP is a formalism that has been initially introduced in the context of problem inversion in image processing. It relies on an untrained neural network whose weights are optimized for each example to inverse. This approach is radically different from classical learning, as it does not require the use of a training set. It is only interested in solving the inverse problem attached to a single example by including explicitly the forward model equations in a deep neural network framework. The neural network (typically a CNN, an encoder-decoder architecture, or a U-Net [29]) takes as initial input a random noise δ . Its structure provides an implicit regularization during the search for a solution, which allows us to find an adequate solution for ill-posed problems. The objective is to find a set of causes X of a phenomenon from the experimental observations of its effects Y . We will denote F the forward model linking X to Y . Typically, X is obtained from Y by solving an optimization process in the whole space of X , which consists of finding the element that best corresponds to the observed effects, once treated by the forward model F . This formulation is often insufficient for the resolution of ill-posed problems and the introduction of additional regularization terms is necessary to constrain the possible X values. However, these regularization terms are often ad hoc and may be insufficient for solving ill-posed problems [15]. The interest of DIP networks lies in their capability of introducing another type of regularization based on the weights of the network itself.

DIP approaches have been first employed for solving well-posed problems such as denoising or inpainting problems. They were later extended to solve ill-posed inverse problems where they provide better results than classical methods on both simulation and experimental data [16].

DIP is hence a good candidate for FPM reconstruction. Indeed, we have implemented a DIP approach with a large number of LEDs ($\Gamma > 60\%$) and obtained a good FPM reconstruction in accordance to [30]. Nevertheless, our first experiments revealed that the quality of reconstructed images is rather poor when a limited number of LEDs is employed (i.e., low overlap), similar to EPRY reconstruction. For these reasons, we propose to incorporate an additional learning step in the model prior to the DIP optimization.

In such a situation, the weights Θ_s of the network are optimized for a large number of examples. The aim of this first step is to approximate the inverse function itself (from various examples) rather than solving the inverse problem for one specific example. After the convergence of this learning step, a second step is undertaken. There, the inverse function (through the weights Θ_s) is solved specifically for the considered data that needs to be inverted. The approximated inverse function obtained from the learning step is used to initialize the optimization step (Figure 2). The same (i.e., unified) cDIP architecture is exploited during these two calculation steps, which ensures a global coherence between the learning step and the optimization one.

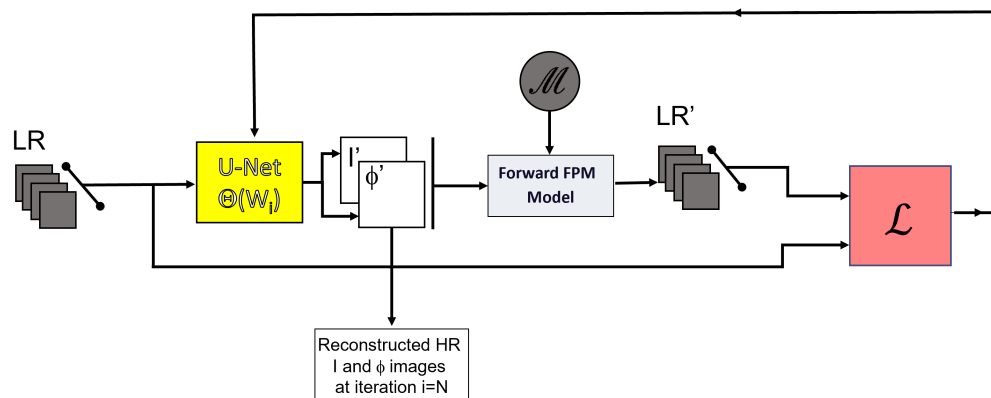


Figure 2. cDIP model: LR designates the stack of FPM low-resolution images measured at camera plan using LEDs \mathcal{M} . Θ represents the parameters of the U-Net that are optimized after each iteration of backpropagation calculations. I' and Φ' represent the HR images reconstructed with U-Net at each iteration. The reconstructed images I, Φ are obtained when the model has converged. LR' is the stack of low-resolution images calculated by the forward model. The loss function \mathcal{L} is defined in Equation (6).

3.1. Reconstruction via cDIP

The architecture of the DIP model is shown in (Figure 2). The model is composed of two blocks, namely a U-Net and a forward FPM block. The U-Net takes as input the complete stack of experimental Low Resolution (LR) images captured with LEDs of \mathcal{M}_A . Its output is a couple of intensity and phase images or equivalently the researched function $\widehat{\mathcal{T}}$ of Equation (5) with $\widehat{\mathcal{T}} = F[I' e^{j\Phi'}]$. The role of the U-Net is to extract information from the images presented at its input. It provides at its output the optimal approximation of the researched solution by minimizing Equation (6). The U-Net performs different convolutions and pooling operations on several layers and its precise architecture is detailed in Appendix A. Furthermore, the guessed solution $\widehat{\mathcal{T}}$ that is updated at each epoch is passed to the forward FPM model. This block is in charge of calculating the raw images attached to the output of the U-Net, using Equation (5). To find the researched solution, one then needs to minimize the MSE between these images and the ones that have been experimentally captured. The resulting loss can be written as

$$\mathcal{L} = \sum_{i=1}^{\mathcal{N}} \left| \sqrt{|F^{-1}[F[I' e^{j\Phi'}](k_x, k_y) \cdot \widehat{C}(k_x - k_x^i, k_y - k_y^i)]|} - \sqrt{I^i(x, y)} \right|^2 \quad (7)$$

where F and F^{-1} are, respectively, the Fourier and inverse Fourier transforms, I' and Φ' represent the CNN's intensity and phase output, \widehat{C} is the Fourier transform of the point spread function, $k_x^{(i)}$ and $k_y^{(i)}$ stand for the projections of the k-vector along the x- and y-axes, corresponding to the i th LED illumination, and $I^i(x, y)$ designates the low-resolution images.

The introduction of the equations describing the microscope image formation in the loss function (Equation (7)) permits us to consider the model as physics-informed in accordance with [19].

In practice, the weights of the network are initialized randomly and optimized through an iterative process aiming at reducing the loss thanks to gradient descent. The process is stopped when the stability of the loss function is obtained.

As explained before, we have also implemented a scheme that relies on a learning dataset as is described in Section 3.2.

3.2. U-Net Learning Scheme for cDIP Initialization (Step 1)

The forward electromagnetic model F attached to Equation (5) is applied to each HR image (I, Φ) of a learning dataset set (LDS) leading to numerous stacks of LR images formed at camera plan for the different illuminations used (see Figure 3). The size of each stack depends on the number of LEDs considered. More precisely, at this stage, we apply F considering the $\mathcal{N} = 130$ LEDs composing \mathcal{M} . We split the resulting LR images into two sub-stacks: those related to the LEDs of \mathcal{M}_A ($\mathcal{N}_A = 37$ images) and to \mathcal{M}_B ($\mathcal{N}_B = 93$ images), respectively.

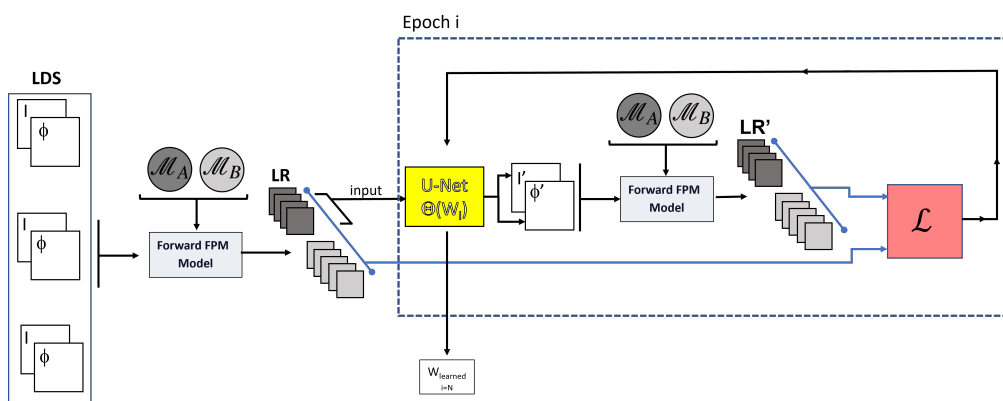


Figure 3. cDIP-L model (U-Net learning model). LDS: learning dataset composed with many intensity I and phase Φ images couples. The forward FPM model is used to calculate low-resolution images formed at the camera plan. \mathcal{M}_A and \mathcal{M}_B represent the splitting of the LED matrix into two groups. LR: stack of low-resolution images formed at camera plan for the different illuminations used. Θ represents the parameters of the U-Net. LR': low-resolution images calculated from the U-Net predicted images I' and Φ' (in dark gray for \mathcal{M}_A , light gray for \mathcal{M}_B). \mathcal{L} : loss function.

Indeed, in this learning stage, we take benefit of the availability of these two sub-stacks (namely the reduced stack and the complementary stack) of LR-images to facilitate the convergence of the network and therefore to obtain a good reconstruction quality.

The \mathcal{N}_A images corresponding to the reduced-led configuration \mathcal{M}_A are the input of a U-Net, which predicts a couple (I', Φ') of HR images. Its parameters Θ_s (the weights) are learned, using a DIP optimization of a loss function, which is the mean squared error between the initial \mathcal{N} images stack and the \mathcal{N} images stack resulting from the forward model applied on (I', Φ') . At each iteration i , a novel couple of images of the learning dataset is considered; in this way, the parameters of the U-Net, the Θ_s , are optimized for a large number of examples, contrary to what occurs in the classic DIP implementation where only one example is considered.

Note that, even if the input of the U-Net is limited to the reduced stack of images, the loss function exploits the complete stack of \mathcal{N} images. In this way, the final reconstructed image benefits from rich information obtained from the original HR images. Furthermore, the loss function does not consider an MSE between (I, Φ) and (I', Φ') as is generally carried out in the literature. Indeed, in the current approach, the loss function results from the comparison of the LR images resulting from the forward model with the complete number of LEDs \mathcal{N} , applied to both couples of HR images. This guarantees that the physical

equations are explicitly solved during the learning step. We denote cDIP-L as the model of step 1 (for the cDIP-Learned model).

3.3. cDIP Reconstruction under Low Overlap Conditions (Step 2)

Once the parameters of the network are learned (after a suitable number of iterations), the model is ready to reconstruct an unknown stack of FPM LR-images acquired under reduced overlap Γ conditions. The reduced stack of \mathcal{N}_A images (related to illuminations \mathcal{M}_A) is presented at the input of the U-Net. As explained before, we have a last step (step 2) in order to refine the solution. This is obtained by solving the forward model that is attached to one single set of LR images only, as detailed in the Section 3.1. It is to be noted that such optimization is performed without any change in the architecture and the same loss function that was introduced in step 1 is exploited. The only difference lies in the fact that the weights of the network $\Theta_{learned}$, obtained after the training phase, are used for the initialization of the U-Net. Also, these weights are now optimized using the reduced stack only since it is the only information available at this stage. The final reconstructed HR image is finally obtained at the output of the network after the convergence of the optimization step.

We note that the interest of our approach is that the models used in the learning and optimization phase are identical. The two steps are hence unified from DNN model point of view. In the following, this model is denoted cDIP-LO (for cDIP learned and optimized model) when the model is initialized with learned weights and cDIP-O when the initialization of DNN weights is random only.

4. Simulation Results and Discussion

The microscope configuration that is employed to evaluate the cDIP-LO reconstruction model consists of a microscope fitted with an objective lens of magnification $4\times$ and $NA = 0.08$. The model is tested with simulations only to evaluate its ultimate performance.

The LED matrix \mathcal{M} composed with 130 LEDs at $\lambda = 525$ nm introduced in Figure 1 is placed at a distance $d = 55$ mm below the sample. The different regions of the sample spectrum that are consequently probed correspond to the many regions already indicated in Figure 1b,c. Such configuration allows us to reach a theoretical super-resolution factor $\gamma = 5$. The pixel pitch of the camera that is simulated is $3.45 \mu\text{m}$. The initialization weights of the cDIP-LO model are obtained using the Learning Data Set LDS (cf. Figure 3) after training the cDIP-L model. LDS is constructed from ImageNet Large Scale Visual Recognition 2019 Challenge (ILSVRC'19) and contains 10,000 complex masks T_j (refer to Equation (1)). Each of them represents a numerical sample used to simulate FPM LR-images. More precisely, T_j is obtained from images taken in the ImageNet database. For each j (j varying from 1 to 10,000), two images of 384×384 pixels (respectively I_j and Φ_j) are randomly extracted from the ImageNet catalog. T_j is then formed with $\sqrt{I_j}e^{j\Phi_j}$. For each T_j , a stack of 130 FPM LR-images simulating the photos that would be acquired by the camera on sample T_j is calculated using a forward model (see Equation (5)) for the different LEDs of the illumination matrix $\mathcal{M} = \mathcal{M}_A \cup \mathcal{M}_B$. The overlap factor Γ is $\sim 60\%$. It is important here to highlight that the LDS that is thereby constructed permits us to simulate a vast variety of samples that would hardly be accessible to experimental FPM measurements. In particular, ImageNet contains a considerable amount of photos with wide variations in their spatial characteristics (contrast, spatial frequencies, etc.). Also, I and Φ functions are uncorrelated by construction. This is usually not the case in real experiments since biological samples absorption and optical thickness are generally closely related to one another. Such LDS could therefore be beneficial for good model generalization. Because of the camera pitch used in the simulations (of $3.45 \mu\text{m}$) and the objective lens characteristics ($4\times NA = 0.08$), the produced LR images are sampled by the camera with an important oversampling factor with respect to Shannon criteria. In our case, each LR image is 128×128 pixels. In complement to LDS, a test dataset TDS is also constituted. It consists of 1000 complex masks that are obtained following the same construction that is used for the LDS. The

images taken from ImageNet are however different than the one used in LDS. For each mask, two stacks of LR images are also calculated using FPM forward model. The first stack is related to LEDs of \mathcal{M}_A , and the second to LEDs of \mathcal{M} .

The TDS is exploited to compare the different reconstruction models (ePIE, cDIP-L, and cDIP-LO). A typical result is presented in Figure 4, where images (a) and (g) correspond to I and Φ HR-images that are reconstructed with the ePIE algorithm using the stack of LR-images relative to LEDs of \mathcal{M} . These two images constitute the targeted (or reference) intensity and phase HR images. Image (b) corresponds to the raw image related to the matrix central LED. There, the native resolution of the objective lens can be appreciated. The other images in Figure 4 are the images that are reconstructed using the LR-images stack relative to LEDs of \mathcal{M}_A only. In particular, images (c) and (h) correspond to ePIE reconstruction, images (d) and (i) to cDIP-L reconstruction, and (f) and (j) to cDIP-LO (intensity and phase). Images (c) and (h) illustrate the typical artifacts that can be obtained with ePIE reconstruction under low overlap conditions. Important crosstalk between the intensity and phase is observed. Also, the phase contrast is highly degraded. It is clear that the quality of these images is too low to consider their exploitation for real applications (with low Γ). In contrast, I and Φ images obtained with cDIP-L are of improved perceptual quality both for intensity and phase. However, artifacts are still noticeable. Some of them are pointed out with arrows in Figure 4. For example, arrow A corresponds to degradations in the high frequencies that are mostly lost. Arrow B corresponds to an incomplete separation between phase and intensity image (crosstalk). Although cDIP-L permits us to recover phase information with benefits as compared to ePIE, the final resolution of reconstructed images is questionable: the intensity image resolution is almost comparable to the raw intensity image relative to central LED illumination. This is consistent with state-of-the-art published results [11], also relying on trained systems. This seems to be inherent to models that are only statistically estimated. Interestingly, the situation is much different with the cDIP-LO model. The differences between reconstructed and reference images are too low to be visually appreciable.

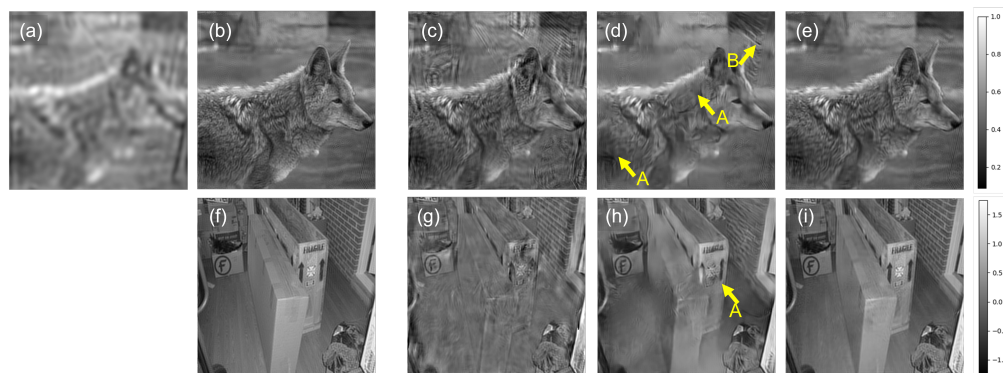


Figure 4. Comparison of reconstruction results obtained with illuminations with an overlap $\sim 10\%$. Images (a–e) correspond to intensity images, whereas images (f–i) to phase images. (a) Raw image acquired with central LED, (b) reference image, (c) ePIE reconstruction (using LEDs \mathcal{M}_A), (d) cDIP-L reconstruction after training (using LEDs \mathcal{M}_A), (e) cDIP-LO reconstruction after training and optimization (using LEDs \mathcal{M}_A). (f) corresponds to reference phase, and (g–i) to reconstructed phase with e-PIE, cDIP-L and cDIP-LO, respectively. The arrows indicate different types of artifacts (such as crosstalk or resolution degradation). References images (b,f) are obtained from ePIE reconstruction using all stack of images (LEDs \mathcal{M}).

This improvement brought by cDIP-LO neural reconstruction is further confirmed by evaluating its performances quantitatively with the error function \mathcal{L} that is calculated over the complete TDS dataset. The traditional SSIM and PSNR metrics are evaluated for reconstructed images with respect to reference images for intensity and phase. The results are indicated in Table 1 where the mean and standard deviation of \mathcal{L} , SSIM, and PSNR

are calculated on the whole TDS. The ePIE column designates ePIE reconstruction using LEDs of \mathcal{M}_A . Unambiguously, the cDIP-LO algorithm leads to better image reconstruction. In particular, the loss evaluated for ePIE and cDIP-L is almost comparable, whereas it is 10 times lower for cDIP-LO. The similarity metric SSIM is close to 1 for cDIP-LO (as desired) and much higher than cDIP-L and ePIE SSIM. The PSNR also reveals an important increase for cDIP-LO compared to cDIP-L and ePIE (above 10 dB).

It is hence clear that the model cDIP-LO improves substantially the reconstruction performances. We attribute such improvement to the fact that the initialization of cDIP-LO weights benefits information that is relative to all the LEDs of \mathcal{M} during the learning step and the fact that cDIP-L reconstruction calculations are further solved using forward model equations. One can also observe from Table 1 that standard deviations of metrics relative to cDIP-LO are low, indicating the ability of this DNN approach to generalize. We recall that TDS has been constituted with various artificial complex masks with no correlation between intensity and phase precisely.

Table 1. \mathcal{L} , SSIM and PSNR metrics evaluated over the complete TDS dataset. The metrics' mean values and standard deviations are indicated. Reconstructions with the different algorithms (ePIE, cDIP-L, and cDIP-LO) are obtained using the stack of LR images obtained with LEDs of \mathcal{M}_A .

	ePIE	cDIP-L	cDIP-LO
$\mathcal{L} \times (10^{-4})$	$\bar{\mu} = 4.43$ $\sigma = 2.85$	$\bar{\mu} = 1.47$ $\sigma = 0.724$	$\bar{\mu} = 0.244$ $\sigma = 0.0941$
$SSIM_{Intensity}$	$\bar{\mu} = 0.569$ $\sigma = 0.143$	$\bar{\mu} = 0.671$ $\sigma = 0.105$	$\bar{\mu} = 0.919$ $\sigma = 0.029$
$SSIM_{Phase}$	$\bar{\mu} = 0.19$ $\sigma = 0.109$	$\bar{\mu} = 0.491$ $\sigma = 0.167$	$\bar{\mu} = 0.915$ $\sigma = 0.06$
$PSNR_{Intensity}$ (dB)	$\bar{\mu} = 17.28$ $\sigma = 4.38$	$\bar{\mu} = 25.03$ $\sigma = 2.68$	$\bar{\mu} = 34.8$ $\sigma = 2.71$
$PSNR_{Phase}$ (dB)	$\bar{\mu} = 18.32$ $\sigma = 3.53$	$\bar{\mu} = 25.05$ $\sigma = 1.86$	$\bar{\mu} = 39.14$ $\sigma = 2.81$

The metrics \mathcal{L} , $SSIM$, and PSNR do not provide information regarding the final delivered image resolution. For this reason, we have undertaken complementary simulations of USAF 1951 intensity target resolution chart reconstruction using LEDs \mathcal{M}_A . Typical results are presented in Figure 5. The first line of images corresponds to the situation where the USAF1951 object is placed at the microscope focal plane. For the second line, the object is out of focus and placed 80 μm above. Figure 5a,b correspond, respectively, to the reference image and the raw image relative to the central LED. The last element that is resolved is indicated with an arrow. The reconstructions obtained with ePIE, cDIP-L, and cDIP-O (i.e., without initialization of the cDIP model) are shown in images Figure 5c–e, respectively. As can be observed, the last resolved element does not change much. That means that no significant improvement in resolution (as compared to native objective lens resolution) is observable with reconstruction, although the global perceptual quality of the reconstructed images is improved. In contrast, cDIP-LO reconstruction (see Figure 5f) reveals a significant change in the position of the last element that is resolved. It permits us to evaluate that the super-resolution factor that is attained is $\gamma = 5$, as is theoretically expected. Note also that no appreciable difference with respect to the reference image is observable.

Although the perceptual quality of the USAF1951 object is minimally a little enhanced as soon as reconstruction is performed (compare Figure 5c–f) to image Figure 5b, the situation is much different when the object is out of focus (second line of images of Figure 5). The images of Figure 5i–l correspond to images reconstructed with digital focus correction (as explained in Section 2). One can note that the quality of images obtained with ePIE, cDIP-L, and cDIP-LO is degraded. In contrast, cDIP-LO demonstrates its ability to reconstruct

correctly the resolution target chart without any degradation. This reconstruction algorithm is hence fully functional and is able to handle correctly the integration of focus correction in the forward model, as is highly desirable for practical situations.

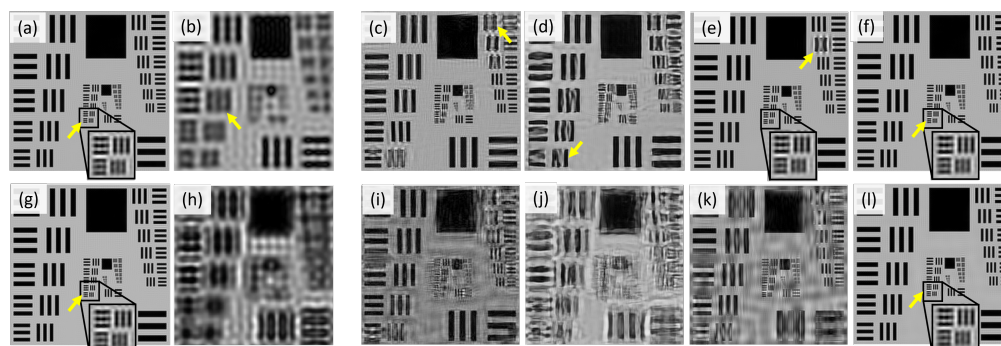


Figure 5. comparison of intensity USAF 1951 target resolution chart reconstructions obtained with illumination \mathcal{M}_A (overlap $\sim 10\%$). For images (a–f), the target resolution chart is placed at the microscope focus plan. (a) Reference, (b) raw image acquired with central LED illumination, (c) ePIE, (d) cDIP-L, (e) cDIP-O, and (f) cDIP-LO reconstructions. For images (g,h), the resolution chart is placed $8\ \mu\text{m}$ above the microscope focus plan. (g) Reference (identical to image (a)), (h) central LED, (i) ePIE, (j) cDIP-L, (k) cDIP-O, and (l) cDIP-LO reconstructions. The reference is calculated using ePIE reconstruction with \mathcal{M} LEDs illumination.

5. Conclusions

In this paper, we have introduced a DNN-based algorithm, namely cDIP-LO DNN, to reconstruct images acquired with FPM under a reduced number of illuminations of the sample (i.e., $\Gamma < 60\%$). Reconstruction properties of such an approach have been compared to the state-of-the-art algorithms ePIE and to the statistically learned DNN model. The simulated results for an FPM configuration with a very low Γ factor ($\Gamma \sim 10\%$) demonstrate the conceptual benefits of cDIP-LO in terms of image quality and resolution. The specificity of cDIP-LO is two-fold: (1) it is a physics-informed DNN, in the sense that the forward image microscope formation model is explicitly introduced in the model to calculate the loss function; the HR reconstructed images are therefore obtained by this model, but their values have never been used directly to calculate the loss, (2) the cDIP-LO model's weights are initialized after a statistical learning step; in this step, the forward model has been purposely introduced in the replacement of any exploitation of a reference solution. This allows the employment of a single and coherent model (or unified DNN architecture) in the learning step and in the optimization step. We note that the learning step is mandatory; the reconstruction with cDIP-LO without initialization of the model leads to unexploitable images.

In future work, the test of cDIP-LO on experimental data will be undertaken. From a practical point of view, we believe that simulated data should be sufficient in the learning step without any need for experimental data. This point will be specifically studied with respect to experimental noise tolerance or possible LEDs position imprecision. The idea of coupling a statistical learning step and a physics-informed DNN optimization is general. It is readily applicable to the classic commercial LED rectangular board. It can also open interesting perspectives for FPM reconstruction under multiplex-coded illumination or for diffraction tomography.

Author Contributions: Conceptualization, L.B., B.D. and Y.G.; Methodology, B.D. and Y.G.; Software, L.B.; Validation, L.B.; Investigation, Y.G.; Writing—original draft, L.B., B.D. and Y.G.; Writing—review and editing, J.K.; Supervision, B.D. and Y.G.; Project administration, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The corresponding author can be contacted for access to the data presented in this study upon request. The data are not publicly available due to restrictions on privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. U-Net Architecture

Among the various state-of-the-art deep network architectures, we selected a U-Net architecture [29]. This choice was guided by the U-Net's demonstrated efficiency in performing tasks such as segmentation, denoising, super-resolution, etc. The architecture implemented here is very similar to the standard U-Net, which is an encoder–decoder model (Figure A1). This model takes as input a stack of low-resolution (LR) images and produces the reconstructed intensity and phase images as output.

The compression part of the U-Net (left side of Figure A1) comprises three operations, repeated four times, as in the classical U-Net: two successive filter blocks of 3×3 convolution kernels, batch normalization (BN), and a non-linear ReLu activation function, followed by 2×2 max-pooling. At each step, the dimension of the images is reduced by a factor of 2. The expansion part of the U-Net (right side of Figure A1) involves 2×2 upsampling, connections to the left side (via skip-connection and concatenation), two successive filter blocks of 3×3 convolution kernels, batch normalization, and a ReLu activation function. The numbers indicated in each convolution block represent the number of filters. We appended a final upsampling stage with a 3×3 upsampling, followed by two convolution and activation blocks to produce an image that complies with the size of the reconstructed images I and Φ .

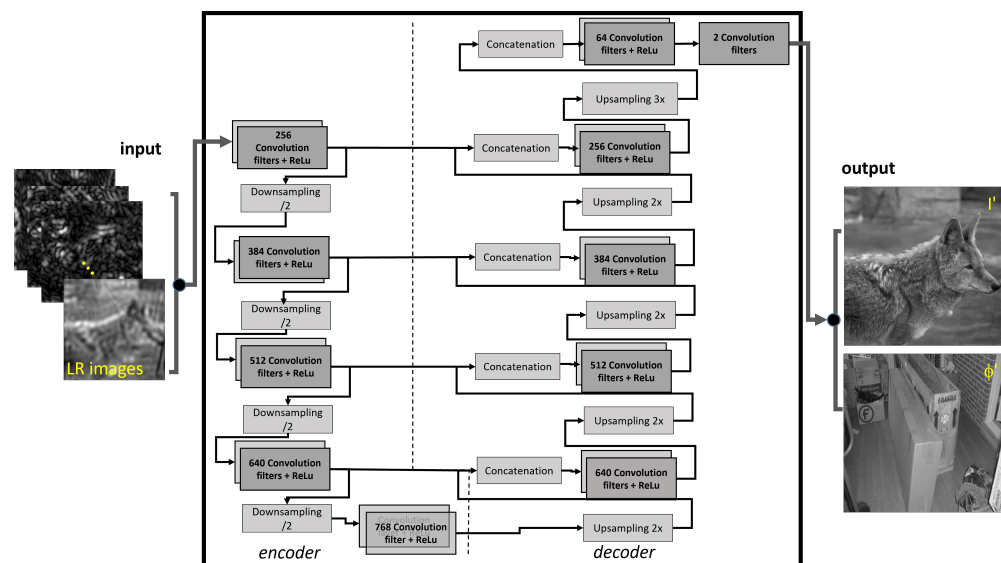


Figure A1. The architecture of the U-Net used to solve the inverse problem for the different schemes.

References

1. Park, J.; Brady, D.J.; Zheng, G.; Tian, L.; Gao, L. Review of bio-optical imaging systems with a high space-bandwidth product. *Adv. Photonics* **2021**, *3*, 044001. <https://doi.org/10.1117/1.ap.3.4.044001>.
2. Shen, C.; Liang, M.; Pan, A.; Yang, C. Non-iterative complex wave-field reconstruction based on Kramers–Kronig relations. *Photonics Res.* **2021**, *9*, 1003–1012. <https://doi.org/10.1364/PRJ.419886>.
3. Tippie, A.E.; Kumar, A.; Fienup, J.R. High-resolution synthetic-aperture digital holography with digital phase and pupil correction. *Opt. Express* **2011**, *19*, 12027–12038. <https://doi.org/10.1364/OE.19.012027>.
4. Ralston, T.S.; Marks, D.L.; Carney, P.S.; Boppart, S.A. Interferometric synthetic aperture microscopy. *Nat. Phys.* **2007**, *3*, 129–134. <https://doi.org/10.1038/nphys514>.
5. Gustafsson, M.G.L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. Short Communication. *J. Microsc.* **2000**, *198*, 82–87. <https://doi.org/10.1046/j.1365-2818.2000.00710.x>.

6. Mico, V.; Zalevsky, Z.; García, J. Synthetic aperture microscopy using off-axis illumination and polarization coding. *Opt. Commun.* **2007**, *276*, 209–217. <https://doi.org/https://doi.org/10.1016/j.optcom.2007.04.020>.
7. Zheng, G.; Horstmeyer, R.; Yang, C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nat. Photonics* **2013**, *7*, 739–745. <https://doi.org/10.1038/nphoton.2013.187>.
8. Pan, A.; Zuo, C.; Yao, B. High-resolution and large field-of-view Fourier ptychographic microscopy and its applications in biomedicine. *Rep. Prog. Phys.* **2020**, *83*, 096101. <https://doi.org/10.1088/1361-6633/aba6f0>.
9. Tian, L.; Liu, Z.; Yeh, L.H.; Chen, M.; Zhong, J.; Waller, L. Computational illumination for high-speed in vitro Fourier ptychographic microscopy. *Optica* **2015**, *2*, 904–911. <https://doi.org/10.1364/OPTICA.2.000904>.
10. Tian, L.; Li, X.; Ramchandran, K.; Waller, L. Multiplexed coded illumination for Fourier Ptychography with an LED array microscope. *Biomed. Opt. Express* **2014**, *5*, 2376–2389. <https://doi.org/10.1364/BOE.5.002376>.
11. Nguyen, T.; Xue, Y.; Li, Y.; Tian, L.; Nehmetallah, G. Deep learning approach for Fourier ptychography microscopy. *Opt. Express* **2018**, *26*, 26470–26484. <https://doi.org/10.1364/OE.26.026470>.
12. Zhang, J.; Xu, T.; Shen, Z.; Qiao, Y.; Zhang, Y. Fourier ptychographic microscopy reconstruction with multiscale deep residual network. *Opt. Express* **2019**, *27*, 8612–8625. <https://doi.org/10.1364/OE.27.008612>.
13. Kappeler, A.; Ghosh, S.; Holloway, J.; Cossairt, O.; Katsaggelos, A. Ptychnet: CNN based fourier ptychography. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 1712–1716. <https://doi.org/10.1109/ICIP.2017.8296574>.
14. Boominathan, L.; Maniparambil, M.; Gupta, H.; Baburajan, R.; Mitra, K. Phase retrieval for Fourier Ptychography under varying amount of measurements. *arXiv* **2018**, arXiv:1805.03593. <https://doi.org/10.48550/ARXIV.1805.03593>.
15. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep Image Prior. *Int. J. Comput. Vis.* **2020**, *128*, 1867–1888. <https://doi.org/10.1007/s11263-020-01303-4>.
16. Dittmer, S.; Kluth, T.; Maass, P.; Baguer, D.O. Regularization by Architecture: A Deep Prior Approach for Inverse Problems. *J. Math. Imaging Vis.* **2019**, *62*, 456–470. <https://doi.org/10.1007/s10851-019-00923-x>.
17. Zhou, K.C.; Horstmeyer, R. Diffraction tomography with a deep image prior. *Opt. Express* **2020**, *28*, 12872–12896. <https://doi.org/10.1364/OE.379200>.
18. Wang, F.; Bian, Y.; Wang, H.; Lyu, M.; Pedrini, G.; Osten, W.; Barbastathis, G.; Situ, G. Phase imaging with an untrained neural network. *Light Sci. Appl.* **2020**, *9*, 77. <https://doi.org/10.1038/s41377-020-0302-3>.
19. Karniadakis, G.E.; Kevrekidis, I.G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440. <https://doi.org/10.1038/s42254-021-00314-5>.
20. Xiang, M.; Pan, A.; Zhao, Y.; Fan, X.; Zhao, H.; Li, C.; Yao, B. Coherent synthetic aperture imaging for visible remote sensing via reflective Fourier ptychography. *Opt. Lett.* **2021**, *46*, 29–32. <https://doi.org/10.1364/OL.409258>.
21. Chen, J.; Wang, A.; Pan, A.; Zheng, G.; Ma, C.; Yao, B. Rapid full-color Fourier ptychographic microscopy via spatially filtered color transfer. *Photonics Res.* **2022**, *10*, 2410–2421. <https://doi.org/10.1364/PRJ.473038>.
22. Fienup, J.R. Phase retrieval algorithms: A comparison. *Appl. Opt.* **1982**, *21*, 2758–2769. <https://doi.org/10.1364/AO.21.002758>.
23. Yang, G.Z.; Dong, B.Z.; Gu, B.Y.; Zhuang, J.Y.; Ersoy, O.K. Gerchberg–Saxton and Yang–Gu algorithms for phase retrieval in a nonunitary transform system: A comparison. *Appl. Opt.* **1994**, *33*, 209–218. <https://doi.org/10.1364/AO.33.000209>.
24. Maiden, A.M.; Rodenburg, J.M.; Humphry, M.J. Optical ptychography: A practical implementation with useful resolution. *Opt. Lett.* **2010**, *35*, 2585–2587. <https://doi.org/10.1364/OL.35.002585>.
25. Ou, X.; Zheng, G.; Yang, C. Embedded pupil function recovery for Fourier ptychographic microscopy. *Opt. Express* **2014**, *22*, 4960–4972. <https://doi.org/10.1364/OE.22.004960>.
26. Sun, J.; Chen, Q.; Zhang, Y.; Zuo, C. Sampling criteria for Fourier ptychographic microscopy in object space and frequency space. *Opt. Express* **2016**, *24*, 15765–15781. <https://doi.org/10.1364/OE.24.015765>.
27. Bunk, O.; Dierolf, M.; Kynde, S.; Johnson, I.; Marti, O.; Pfeiffer, F. Influence of the overlap parameter on the convergence of the ptychographical iterative engine. *Ultramicroscopy* **2008**, *108*, 481–487. <https://doi.org/10.1016/j.ultramic.2007.08.003>.
28. Jiang, S.; Guo, K.; Liao, J.; Zheng, G. Solving Fourier ptychographic imaging problems via neural network modeling and TensorFlow. *Biomed. Opt. Express* **2018**, *9*, 3306–3319. <https://doi.org/10.1364/BOE.9.003306>.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
30. Konda, P.C.; Loetgering, L.; Zhou, K.C.; Xu, S.; Harvey, A.R.; Horstmeyer, R. Fourier ptychography: Current applications and future promises. *Opt. Express* **2020**, *28*, 9603–9630. <https://doi.org/10.1364/OE.386168>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.