OPEN ACCESS
IOP Publishing

Inverse Problems

Inverse Problems **39** (2023) 064005 (23pp) https://doi.org/10.1088/1361-6420/acca72

Dynamic Fourier ptychography with deep spatiotemporal priors

Pakshal Bohra[1,4,*], Thanh-an Pham[2,4], Yuxuan Long[1], Jaejun Yoo[3] and Michael Unser[1]

[1] Biomedical Imaging Group, EPFL, Lausanne, Switzerland
[2] 3D Optical Systems Group, MIT, Cambridge, MA, United States of America
[3] Laboratory of Advanced Imaging Technology, UNIST, Ulsan, Republic of Korea

E-mail: pakshal.bohra@epfl.ch

Received 31 August 2022; revised 12 March 2023
Accepted for publication 5 April 2023
Published 9 May 2023

Abstract

Fourier ptychography (FP) involves the acquisition of several low-resolution intensity images of a sample under varying illumination angles. They are then combined into a high-resolution complex-valued image by solving a phase-retrieval problem. The objective in dynamic FP is to obtain a sequence of high-resolution images of a moving sample. There, the application of standard frame-by-frame reconstruction methods limits the temporal resolution due to the large number of measurements that must be acquired for each frame. In this work instead, we propose a neural-network-based reconstruction framework for dynamic FP. Specifically, each reconstructed image in the sequence is the output of a shared deep convolutional network fed with an input vector that lies on a one-dimensional manifold that encodes time. We then optimize the parameters of the network to fit the acquired measurements. The architecture of the network and the constraints on the input vectors impose a spatiotemporal regularization on the sequence of images. This enables our method to achieve high temporal resolution without compromising the spatial resolution. The proposed framework does not require training data. It also recovers the pupil

[4] These authors contributed equally to this paper.
[*] Author to whom any correspondence should be addressed.

1361-6420/23/064005+23$33.00 © 2023 The Author(s). Published by IOP Publishing Ltd Printed in the UK

function of the microscope. Through numerical experiments, we show that our framework paves the way for high-quality ultrafast FP.

Supplementary material for this article is available online

Keywords: Fourier ptychography, dynamic imaging, regularization, neural networks

(Some figures may appear in colour only in the online journal)

1. Introduction

In Fourier ptychography (FP) [1], hundreds of low-resolution intensity images are acquired by illuminating the object of interest with a coherent light source with varying incidence angles. This task is typically performed using a LED array and a microscope with a low numerical aperture (NA) objective lens, which makes FP a low-cost and label-free imaging modality. The collection of measurements is then algorithmically combined into a high-resolution complex-valued image of the sample over a large field of view. Thus, FP has a high space-bandwidth product.

Building upon the pioneering work of Zheng *et al* [1], the capabilities of FP have been extended in a variety of ways by improving the optical acquisition setup. For instance, in [2, 3], the sequence of illuminations is optimized via an importance metric and neural networks, respectively. Multiplexed FP is introduced in [4], where one illuminates the sample with multiple LEDs and is able to reduce the number of measurements. Further, optimal combinations of LEDs are studied in [3, 5–7].

There have also been several improvements on the computational side for FP. At its core, the reconstruction process involves the solution of a phase-retrieval (PR) problem—the recovery of phase information from intensity measurements. In [1], this task is performed by using the iterative Gerchberg–Saxton (GS) algorithm [8]. As PR is a non-convex problem, the solution obtained by GS depends on the starting point. This problem of initialization is tackled in [9]. In [10–12], PR is formulated as a convex optimization problem with the help of a lifting scheme. However, this elegant approach comes at the cost of a large computational burden. As the acquired measurements are typically corrupted by noise, maximum-likelihood estimation offers an adequate framework for one to incorporate the noise statistics [13]. The resulting optimization problems are solved efficiently by gradient-based or higher-order methods [14, 15]. A thorough comparative study of different methods for PR can be found in [16]. In addition to solving the PR problem, algorithms that include the estimation of the pupil function of the microscope [17] and correction of the LED positions [18, 19] have also been proposed.

While FP has matured into a versatile modality with numerous applications [20], high-quality high-speed imaging remains a challenge. The temporal resolution in FP is inherently limited by the large number of measurements that need to be acquired in order to reconstruct the high-resolution image of the sample. To alleviate this problem, ad hoc acquisition setups [5, 6, 21] have been devised. They allow one to obtain a higher temporal resolution without a significant deterioration of the spatial resolution. Alternatively, there has been a lot of interest in the development of sophisticated computational methods to solve the PR problem with only a few measurements. In such ill-posed scenarios, regularization techniques can be used to incorporate some prior knowledge about the sample of interest. These are typically applied by formulating PR as an optimization problem where the cost functional consists of a data-fidelity term and a regularization term. The data-fidelity term ensures that the solution is consistent with the observed data while the regularization promotes solutions with the desired properties. For example, the popular total-variation (TV) regularization [22] favors piecewise-constant images and has been adapted for FP in several works [23–26]. Group-sparsity-based priors have been successfully deployed in FP as well [27]. An online plug-and-play approach for FP has also been proposed in [28], where sophisticated denoisers such as BM3D [29] are used for (implicit) regularization.

Over the past few years, deep-learning-based methods have yielded impressive results, outperforming the model-based regularized methods in a variety of imaging modalities, especially in ill-posed settings [30, 31]. In the context of FP, deep neural networks have been trained in a supervised manner as nonlinear mappings that take the low-resolution measurements and output the high-resolution image of interest [32–34]. Further, in [35, 36], pre-trained deep generative priors are used to solve the PR problem. For more details regarding FP, we refer the reader to recent comprehensive reviews [20, 37].

In dynamic FP, when it is desired to image a moving sample, the computational methods described above must be applied in a frame-by-frame manner to obtain the sequence of high-resolution images, without accounting for the temporal dependencies in the measurements. Yet, one can decrease the number of measurements required per frame (thus increasing the effective imaging speed) by exploiting the temporal correlations in the sequence of images to be recovered. Based on this idea, the concept of low-rank FP is introduced in [38], where a low-rank constraint is enforced on the matrix formed by stacking the (vectorized) images.

1.1. Contributions

In this work, we propose a novel computational framework for dynamic FP. Inspired by the method developed in [39] for dynamic magnetic resonance imaging, we use a deep neural network (deep prior) to impose a spatiotemporal regularization on the sequence of complex-valued images to be recovered. More specifically, we parameterize each image in the sequence as the output of a single convolutional network corresponding to some fixed latent input vector. These input vectors are chosen to lie on a one-dimensional manifold. The parameters of the network are then optimized such that the generated images collectively fit the acquired measurements under the action of the specified forward model. The architecture of the generative network imposes an implicit spatial prior on the images while the constraints on the input latent vectors allow the network to associate their proximity with temporal variations in the sequence. Our method does not require any training data. It also estimates the pupil function together with the complex-valued images, which means it can be readily applied for different settings. We assess the performance of our framework on simulated data with a single measured low-resolution image per reconstructed frame and show that it paves the way for high-quality ultrafast FP.

The paper is organized as follows. In section 2, we describe a continuous-domain physical model for FP along with its computationally efficient discretization. We present the proposed reconstruction framework in section 3 and the experimental results in section 4.

2. Physical model

In this section, we first formulate the physical model that relates the acquired measurements and the sample of interest in the continuous domain. Then, we present a discretized version of the forward model that can be implemented in a computationally efficient manner.

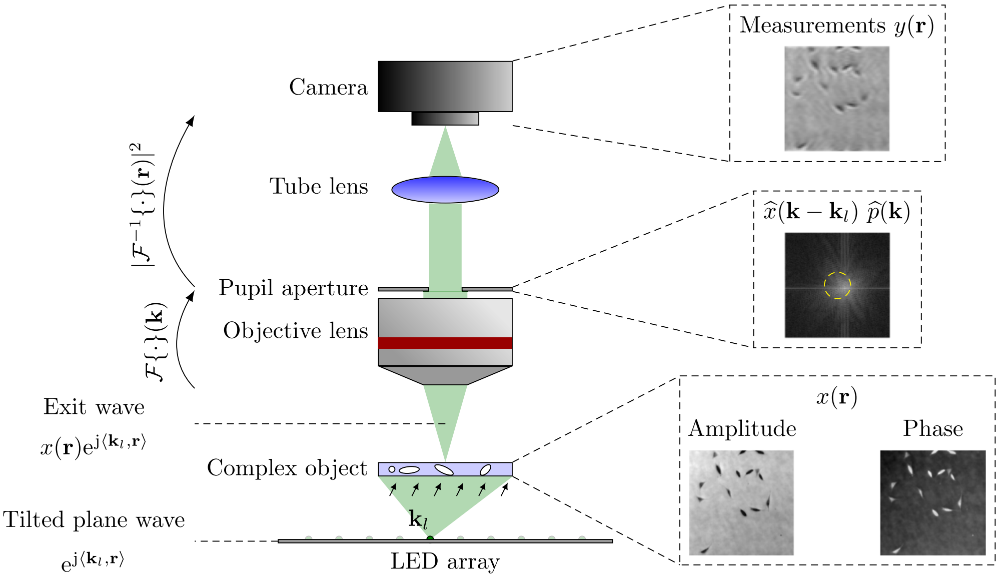

Measurements $y(\mathbf{r})$
Camera
Tube lens
$|\mathcal{F}^{-1}\{\cdot\}(\mathbf{r})|^2$
$\widehat{x}(\mathbf{k}-\mathbf{k}_l)\ \widehat{p}(\mathbf{k})$
Pupil aperture
Objective lens
$\mathcal{F}\{\cdot\}(\mathbf{k})$
Exit wave
$x(\mathbf{r})\mathrm{e}^{\mathrm{j}\langle\mathbf{k}_l,\mathbf{r}\rangle}$
$x(\mathbf{r})$
Amplitude
Phase
Complex object
Tilted plane wave
$\mathrm{e}^{\mathrm{j}\langle\mathbf{k}_l,\mathbf{r}\rangle}$
$\mathbf{k}_l$
LED array

Figure 1. Acquisition setup of Fourier ptychography.

2.1. Continuous-domain formulation

The optical system in FP usually involves an array of L LEDs (see figure 1), where the lth LED illuminates the specimen with a tilted plane wave with wave vector $\mathbf{k}_l \in \mathbb{R}^2$ ($l \in \mathcal{L} = \{1,2,\ldots,L\}$) and wavelength $\lambda > 0$. In this work, we consider the case where only one LED is turned on for each measured image. However, our framework is also compatible with more sophisticated acquisition settings [4].

We model the sample of interest as a 2D complex object, which is a valid assumption for thin samples. Therefore, we can represent the moving sample as a complex-valued function $x : \Omega_\mathrm{X} \times \mathbb{R}_{\geqslant 0} \to \mathbb{C}$, where $\Omega_\mathrm{X} \subset \mathbb{R}^2$ includes the region of interest of the sample. Let $\{t_q\}_{q=1}^{Q}$ be the uniformly-spaced timestamps, with spacing Δ_t, at which we are interested in observing the sample. We assume that the sample moves very slowly in the intervals $\{T_q = [t_q - \Delta_\mathrm{t}/2, t_q + \Delta_\mathrm{t}/2]\}_{q=1}^{Q}$. Thus, during T_q, we can acquire multiple measurements $\{y_{q,w} : \Omega_\mathrm{Y} \to \mathbb{R}\}_{w=1}^{W}$, where $W \leqslant L$ and where $\Omega_\mathrm{Y} \subset \mathbb{R}^2$ includes the support of the measurement, of the object $x(\cdot, t_q)$. Here, the tradeoff between the temporal resolution and the spatial resolution can be understood in terms of Δ_t and W: a small value of Δ_t (high temporal resolution) implies a small value of W, which yields a low spatial resolution.

Let $\mathcal{I}_q \subset \mathcal{L}$, where $q \in \{1,2,\ldots,Q\}$, be the set of LEDs that are switched on during T_q; the cardinality of this set is $|\mathcal{I}_q| = W$. Further, for $w \in \{1,2,\ldots,W\}$, we introduce $l_{q,w} = \mathcal{I}_q(w) \in \mathcal{L}$ to denote the wth entry of $\mathcal{I}_q$. The measurement image $y_{q,w}$ is obtained when $x(\cdot, t_q)$ is illuminated by the $l_{q,w}$th LED with the tilted plane wave $\mathbf{r} \mapsto \mathrm{e}^{\mathrm{j}\langle \mathbf{k}_{l_{q,w}}, \mathbf{r}\rangle}$. As mentioned in [4, 16], it is given by

$$\begin{aligned} y_{q,w}(\mathbf{r}) &= \left|\mathcal{F}^{-1}\left\{\widehat{p}(\mathbf{k})\mathcal{F}\left\{x(\cdot,t_q)\mathrm{e}^{\mathrm{j}\langle \mathbf{k}_{l_{q,w}},\cdot\rangle}\right\}(\mathbf{k})\right\}(\mathbf{r})\right|^2 \\ &= \left|\mathcal{F}^{-1}\left\{\widehat{p}(\mathbf{k})\widehat{x}(\mathbf{k}-\mathbf{k}_{l_{q,w}},t_q)\right\}(\mathbf{r})\right|^2. \end{aligned} \tag{1}$$

Here, the operators $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse, respectively, $\mathbf{k} \in \mathbb{R}^2$ is the 2D spatial frequency variable, and the quantity $\widehat{x}(\mathbf{k}, t_q)$ denotes the Fourier transform of $x(\mathbf{r}, t_q)$. The pupil function[5] $\widehat{p} : \mathbb{R}^2 \to \mathbb{C}$ models the pupil aperture and is compactly supported on a disk of radius $2\pi \frac{\mathrm{NA}}{\lambda}$, where NA is the numerical aperture of the system, thus cutting off high frequencies.

2.2. Camera sampling

In practice, the camera in the acquisition setup samples $y_{q,w}$ on a uniform grid with stepsize Δ and records a discrete image $\widetilde{\mathbf{y}}^{\mathrm{im}}_{q,w}$ of size[6] $(M \times M)$ such that

$$\widetilde{\mathbf{y}}^{\mathrm{im}}_{q,w} = \mathrm{Noise}(\mathbf{y}^{\mathrm{im}}_{q,w}), \tag{2}$$

where the $(M \times M)$ image $\mathbf{y}^{\mathrm{im}}_{q,w}$ is the sampled version of $y_{q,w}$ given by

$$\mathbf{y}^{\mathrm{im}}_{q,w}[m_1, m_2] = y_{q,w}\Big((m_1 - M/2)\Delta, (m_2 - M/2)\Delta\Big) \tag{3}$$

for $m_1 = 0, \ldots, (M-1)$ and $m_2 = 0, \ldots, (M-1)$, and the operator $\mathrm{Noise}(\cdot)$ models the corruption of $\mathbf{y}^{\mathrm{im}}_{q,w}$ by noise. Consider the quantity

$$u_{q,w}(\mathbf{r}) = \mathcal{F}^{-1}\left\{\widehat{p}(\mathbf{k})\mathcal{F}\left\{x(\cdot, t_q)\mathrm{e}^{\mathrm{j}\langle \mathbf{k}_{t_{q,w}}, \cdot\rangle}\right\}(\mathbf{k})\right\}(\mathbf{r}). \tag{4}$$

Due to the compact support of the pupil function $\widehat{p}$, the maximum angular frequency of $u_{q,w}$ is $2\pi \frac{\mathrm{NA}}{\lambda}$. Note that the Fourier transform of $y_{q,w}$ can be written as

$$\mathcal{F}\{y_{q,w}\}(\mathbf{k}) = \mathcal{F}\left\{|u_{q,w}|^2\right\}(\mathbf{k}) = \left(\overline{\widehat{u}^{\vee}_{q,w}} * \widehat{u}_{q,w}\right)(\mathbf{k}), \tag{5}$$

where $\overline{\widehat{u}^{\vee}_{q,w}}$ denotes the complex conjugate of $\widehat{u}^{\vee}_{q,w}$ which is given by $\widehat{u}^{\vee}_{q,w}(\mathbf{k}) = \widehat{u}_{q,w}(-\mathbf{k})$, and $*$ denotes the convolution operation. Thus, the maximum angular frequency of $y_{q,w}$ is $\frac{4\pi\mathrm{NA}}{\lambda}$. Consequently, the Nyquist criterion dictates that the sampling step Δ of the camera should satisfy

$$\Delta \leqslant \frac{\lambda}{4\mathrm{NA}}. \tag{6}$$

2.3. Discretized forward model

In this work, we obtain a discrete version $\mathbf{x}^{\mathrm{im}}_q$ of $x(\mathbf{r}, t_q)$ by sampling it on a uniform $(N \times N)$ grid with pixel-size Δ_{r}, as

$$\mathbf{x}^{\mathrm{im}}_q[n_1, n_2] = x\Big((n_1 - N/2)\Delta_{\mathrm{r}}, (n_2 - N/2)\Delta_{\mathrm{r}}, t_q\Big) \tag{7}$$

for $n_1 = 0, \ldots, (N-1)$ and $n_2 = 0, \ldots, (N-1)$. The image size is given by $N = r_{\mathrm{p}}M$, where $r_{\mathrm{p}} = \Delta/\Delta_{\mathrm{r}} \in \mathbb{N}$ is the upsampling factor. Now, consider the 2D discrete Fourier transform (DFT) of $\mathbf{x}^{\mathrm{im}}_q$. The corresponding pixel size in the Fourier domain (or angular frequency resolution) is $\Delta_{\mathrm{k}} = 2\pi/N\Delta_{\mathrm{r}}$. Thus, we discretize the pupil function such that

$$\widehat{\mathbf{p}}^{\mathrm{im}}[k_1, k_2] = \widehat{p}\Big((k_1 - M/2)\Delta_{\mathrm{k}}, (k_2 - M/2)\Delta_{\mathrm{k}}\Big) \tag{8}$$

for $k_1 = 0, \ldots, (M-1)$ and $k_2 = 0, \ldots, (M-1)$. Note that the choice of Δ and Δ_{k} ensures that the support of the pupil function lies within the $(M \times M)$ sampling grid for $\widehat{p}$. Moreover,

[5] The pupil function $\widehat{p}$ is described directly in the Fourier domain.

[6] We consider square even-sized images for the sake of simplicity.

in our discretization scheme, we assume that the wave vector $\mathbf{k}_{l_{q,w}}$ can be written as $\mathbf{k}_{l_{q,w}} = (b_{l_{q,w},1}\Delta_{\mathrm{k}}, b_{l_{q,w},2}\Delta_{\mathrm{k}})$, where $b_{l_{q,w},1}, b_{l_{q,w},2} \in \mathbb{Z}$.

We now introduce some additional notations to specify the discrete forward model. Let $\widetilde{\mathbf{y}}_{q,w} \in \mathbb{R}^{M^2}$, $\mathbf{y}_{q,w} \in \mathbb{R}^{M^2}$, $\mathbf{x}_q \in \mathbb{C}^{N^2}$, and $\widehat{\mathbf{p}} \in \mathbb{C}^{M^2}$ be the vectorized versions of $\widetilde{\mathbf{y}}^{\mathrm{im}}_{q,w}$, $\mathbf{y}^{\mathrm{im}}_{q,w}$, $\mathbf{x}^{\mathrm{im}}_q$, and $\widehat{\mathbf{p}}^{\mathrm{im}}$, respectively. Then, let $\mathbf{F}_Q, \mathbf{F}_Q^{-1} \in \mathbb{C}^{Q^2 \times Q^2}$ be matrices that represent the 2D DFT and its inverse of a $(Q \times Q)$ image, respectively. Next, we define $\mathbf{diag}(\widehat{\mathbf{p}}) \in \mathbb{C}^{M^2 \times M^2}$ to be a diagonal matrix whose entries are the values in $\widehat{\mathbf{p}}$. Finally, $\mathbf{C}_{\mathbf{k}_{l_{q,w}}}$ is a boolean matrix that restricts an N^2-dimensional vector to an M^2-dimensional vector depending on the illumination wave vector $\mathbf{k}_{l_{q,w}}$.

Proposition 1. *The discrete counterpart of (1) can be computed as*

$$\mathbf{y}_{q,w} = |\mathbf{H}_{l_{q,w}}\mathbf{x}_q|^2 = \left|\frac{4\pi^2}{r_{\mathrm{p}}^2}\mathbf{F}_M^{-1}\mathbf{diag}(\widehat{\mathbf{p}})\mathbf{C}_{\mathbf{k}_{l_{q,w}}}\mathbf{F}_N\mathbf{x}_q\right|^2. \tag{9}$$

Proof. Consider the quantity

$$\begin{aligned} u_{q,w}(\mathbf{r}) &= \mathcal{F}^{-1}\left\{\widehat{p}(\mathbf{k})\mathcal{F}\left\{x(\cdot,t_q)\mathrm{e}^{\mathrm{j}\langle \mathbf{k}_{l_{q,w}},\cdot\rangle}\right\}(\mathbf{k})\right\}(\mathbf{r}) \\ &= \int_{\mathbb{R}^2}\widehat{p}(\mathbf{k})\mathrm{e}^{\mathrm{j}\langle\mathbf{k},\mathbf{r}\rangle}\left(\int_{\mathbb{R}^2}x(\mathbf{s},t_q)\mathrm{e}^{-\mathrm{j}\langle\mathbf{k}-\mathbf{k}_{l_{q,w}},\mathbf{s}\rangle}\mathrm{d}\mathbf{s}\right)\mathrm{d}\mathbf{k}. \end{aligned} \tag{10}$$

We discretize the integrals in (10) using Riemann sums. A step-size Δ_{k} is used for the integral with respect to $\mathbf{k}$ and a step-size Δ_{r} is used for the integral with respect to $\mathbf{s}$. The samples $\mathbf{u}^{\mathrm{im}}_{q,w}[m_1,m_2] = u_{q,w}\big((m_1 - M/2)\Delta, (m_2 - M/2)\Delta\big)$ for $m_1 = 0,\ldots,(M-1)$ and $m_2 = 0,\ldots,(M-1)$, are then given by

$$\begin{aligned} \mathbf{u}^{\mathrm{im}}_{q,w}[m_1,m_2] = \ &(\Delta_{\mathrm{k}}\Delta_{\mathrm{r}})^2\sum_{k_1=0}^{M-1}\sum_{k_2=0}^{M-1}\Bigg(\underbrace{\widehat{p}\big((k_1-M/2)\Delta_{\mathrm{k}},(k_2-M/2)\Delta_{\mathrm{k}}\big)}_{\widehat{\mathbf{p}}^{\mathrm{im}}[k_1,k_2]} \\ &\times \mathrm{e}^{\mathrm{j}(k_1-M/2)(m_1-M/2)\Delta_{\mathrm{k}}\Delta}\ \mathrm{e}^{\mathrm{j}(k_2-M/2)(m_2-M/2)\Delta_{\mathrm{k}}\Delta}\ \mathbf{a}_{q,w}[k_1,k_2]\Bigg), \end{aligned} \tag{11}$$

where

$$\begin{aligned} \mathbf{a}_{q,w}[k_1,k_2] = &\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}\Bigg(\underbrace{x\big((n_1-N/2)\Delta_{\mathrm{r}},(n_2-N/2)\Delta_{\mathrm{r}},t_q\big)}_{\mathbf{x}^{\mathrm{im}}_q[n_1,n_2]} \\ &\times \mathrm{e}^{-\mathrm{j}(k_1-b_{l_{q,w},1}-M/2)(n_1-N/2)\Delta_{\mathrm{k}}\Delta_{\mathrm{r}}}\ \mathrm{e}^{-\mathrm{j}(k_2-b_{l_{q,w},2}-M/2)(n_2-N/2)\Delta_{\mathrm{k}}\Delta_{\mathrm{r}}}\Bigg). \end{aligned} \tag{12}$$

The limits in the sums in (11) and (12) are dictated by the supports of $\widehat{p}$ and $x(\mathbf{r},t_q)$, respectively. By rearranging some terms and using the fact that $\Delta_{\mathrm{k}}\Delta_{\mathrm{r}} = 2\pi/N$, we rewrite (12) as

$$\begin{aligned} \mathbf{a}_{q,w}[k_1,k_2] = &\underbrace{\left(\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}\mathbf{x}^{\mathrm{im}}_q[n_1,n_2]\ \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}(k_1-b_{l_{q,w},1}-M/2)n_1}\ \mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}(k_2-b_{l_{q,w},2}-M/2)n_2}\right)}_{\widehat{\mathbf{x}}^{\mathrm{im}}_q[k_1-b_{l_{q,w},1}-M/2,k_2-b_{l_{q,w},2}-M/2]} \\ &\times \mathrm{e}^{\mathrm{j}\pi(k_1-b_{l_{q,w},1}-M/2)}\ \mathrm{e}^{\mathrm{j}\pi(k_2-b_{l_{q,w},2}-M/2)}, \end{aligned} \tag{13}$$

where $\widehat{\mathbf{x}}^{\mathrm{im}}_q$ is the (N,N)-point DFT of $\mathbf{x}^{\mathrm{im}}_q$ and the shifts in the DFT are applied in a circular manner. On plugging (13) into (11), we get that

$$\begin{aligned}\mathbf{u}^{\mathrm{im}}_{q,w}[m_1,m_2] = (2\pi/N)^2 \sum_{k_1=0}^{M-1}\sum_{k_2=0}^{M-1} \Bigg(& \widehat{\mathbf{p}}^{\mathrm{im}}[k_1,k_2]\, \widehat{\mathbf{x}}^{\mathrm{im}}_q[k_1 - b_{l_{q,w},1} - M/2, k_2 - b_{l_{q,w},2} - M/2] \\ & \times \mathrm{e}^{\mathrm{j}(k_1-M/2)(m_1-M/2)\Delta_{\mathrm{k}}\Delta}\, \mathrm{e}^{\mathrm{j}(k_2-M/2)(m_2-M/2)\Delta_{\mathrm{k}}\Delta} \\ & \times \mathrm{e}^{\mathrm{j}\pi\,(k_1-b_{l_{q,w},1}-M/2)}\, \mathrm{e}^{\mathrm{j}\pi\,(k_2-b_{l_{q,w},2}-M/2)}\Bigg). \end{aligned} \tag{14}$$

Next, we group all the exponential terms involving k_1 and k_2 and use $\Delta_{\mathrm{k}}\Delta = 2\pi/M$ to obtain that

$$\begin{aligned}\mathbf{u}^{\mathrm{im}}_{q,w}[m_1,m_2] = (2\pi/N)^2 \Bigg(& \sum_{k_1=0}^{M-1}\sum_{k_2=0}^{M-1} \widehat{\mathbf{p}}^{\mathrm{im}}[k_1,k_2]\, \widehat{\mathbf{x}}^{\mathrm{im}}_q[k_1 - b_{l_{q,w},1} - M/2, k_2 - b_{l_{q,w},2} - M/2] \\ & \times \mathrm{e}^{\mathrm{j}\frac{2\pi}{M}k_1 m_1}\, \mathrm{e}^{\mathrm{j}\frac{2\pi}{M}k_2 m_2}\Bigg) \times \mathrm{e}^{-\mathrm{j}\pi\,(m_1+b_{l_{q,w},1})}\, \mathrm{e}^{-\mathrm{j}\pi\,(m_2+b_{l_{q,w},2})}. \end{aligned} \tag{15}$$

Let $\mathbf{g}^{\mathrm{im}}_{q,w}$ be the (M,M)-point inverse discrete Fourier transform (IDFT) of $\widehat{\mathbf{g}}^{\mathrm{im}}_{q,w}[k_1,k_2] = \widehat{\mathbf{p}}^{\mathrm{im}}[k_1,k_2]\, \widehat{\mathbf{x}}^{\mathrm{im}}_q[k_1 - b_{l_{q,w},1} - M/2, k_2 - b_{l_{q,w},2} - M/2]$. Then, the discrete measurements can be expressed as

$$\mathbf{y}^{\mathrm{im}}_{q,w}[m_1,m_2] = \left|\mathbf{u}^{\mathrm{im}}_{q,w}[m_1,m_2]\right|^2 = \left|(4\pi^2/r_{\mathrm{p}}^2)\, \mathbf{g}^{\mathrm{im}}_{q,w}[m_1,m_2]\right|^2. \tag{16}$$

Note that the computation of $\mathbf{g}^{\mathrm{im}}_{q,w}$ involves taking the (N,N)-point DFT of $\mathbf{x}^{\mathrm{im}}_q$, (circularly) shifting it according to the wave vector $\mathbf{k}_{l_{q,w}}$, restricting the shifted DFT to an $(M\times M)$ image, performing pointwise multiplication with $\widehat{\mathbf{p}}^{\mathrm{im}}$, and then taking the (M,M)-point IDFT. This allows us to write (16) in vectorized form as in the right-hand side of (9). □

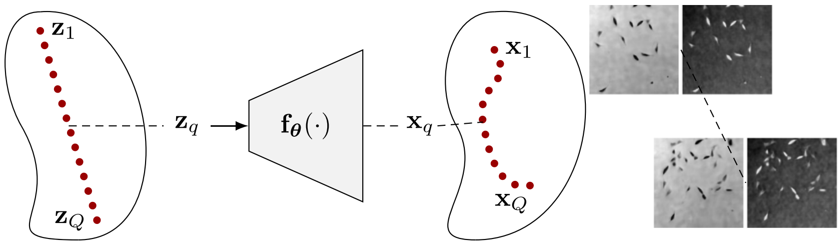
While the discrete forward model (9) has previously been used in works such as [16], to the best of our knowledge, a systematic derivation of (9) from the continuous model (1) has not been presented in the literature.

3. Reconstruction framework

The goal in dynamic FP is to reconstruct the images $\{\mathbf{x}_q \in \mathbb{C}^{N^2}\}_{q=1}^{Q}$ from the recorded measurements $\{\{\widetilde{\mathbf{y}}_{q,w} \in \mathbb{R}^{M^2}\}_{w=1}^{W}\}_{q=1}^{Q}$. We first present our neural-network-based framework for the case of a well-characterized pupil function. Then, we describe a way to incorporate the recovery of the pupil function into our reconstruction algorithm.

3.1. Deep spatiotemporal priors

The concept of using untrained convolutional neural networks (CNNs) as regularization for solving inverse problems was first introduced in [40] under the name ‘deep image prior’ (DIP). There, the image of interest is represented as the output of a CNN with adjustable



Optimization problem: $\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^P} \sum_{q=1}^{Q}\sum_{w=1}^{W} \mathcal{D}\big(\tilde{\mathbf{y}}_{q,w}, |\mathbf{H}_{l_{q,w}}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)|^2\big)$

Figure 2. Spatiotemporal regularization using the generative neural network $\mathbf{f}_{\boldsymbol{\theta}}$.

parameters and some fixed input. It is then shown that the fitting of the network to the measurements yields high-quality image reconstructions for several applications such as denoising, superresolution, and inpainting. This is attributed to the observation that CNNs have a remarkable tendency to favor natural-looking images ('good' solutions) over noisy ones ('bad' solutions). In some scenarios, DIP is deployed with early stopping as deep networks have the capacity to fit noise. In other words, there is a point beyond which running the optimization process for more iterations degrades the quality of the reconstruction. Thus, the architecture of a CNN (and the optimization procedure) can be used as an implicit prior for natural images.

In our reconstruction framework, we propose to use an extended version of DIP to impose spatiotemporal regularization on the sequence of images. We parameterize each of the Q images as the output of a single CNN $\mathbf{f}_{\boldsymbol{\theta}}: \mathbb{R}^{N_z^2} \to \mathbb{C}^{N^2}$, with adjustable parameters $\boldsymbol{\theta} \in \mathbb{R}^P$, applied to some fixed input latent vector $\mathbf{z}_q \in \mathbb{R}^{N_z^2}$, $q = 1, \ldots, Q$. We choose these latent vectors such that they lie on a straight line, in accordance with

$$\mathbf{z}_q = \mathbf{z}_1 + \frac{q-1}{Q-1}(\mathbf{z}_Q - \mathbf{z}_1), \qquad q = 1, \ldots, Q, \tag{17}$$

where the end-points $\mathbf{z}_1, \mathbf{z}_Q$ are fixed beforehand (for example, by drawing two samples from some multivariate probability distribution). We then optimize the parameters of the network to fit the measurements according to

$$\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^P} \sum_{q=1}^{Q}\sum_{w=1}^{W} \mathcal{D}\big(\tilde{\mathbf{y}}_{q,w}, |\mathbf{H}_{l_{q,w}}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)|^2\big), \tag{18}$$

where $\mathcal{D}: \mathbb{R}^{M^2} \times \mathbb{R}^{M^2} \to \mathbb{R}_+$ is the data-fidelity term and the reconstructed sequence is $\{\mathbf{x}_q^*\}_{q=1}^{Q} = \{\mathbf{f}_{\boldsymbol{\theta}^*}(\mathbf{z}_q)\}_{q=1}^{Q}$. The rationale behind our choice of the latent vectors is to allow the CNN to associate the spatial proximity between them with the temporal proximity of the images. In this manner, the architecture of the network imposes spatial regularization while the use of a shared network for all images and the design of the latent space impose temporal regularization. A schematic illustration of our framework is given in figure 2.

Algorithm 1. Initialization of network parameters.

```
Input: Low-quality reconstructions $\{\widetilde{\mathbf{x}}_q\}_{q=1}^{Q}$, latent vectors $\{\mathbf{z}_q\}_{q=1}^{Q}$, batch size $B_Q$, tolerance $\epsilon_{\mathrm{tol}}$, maximum number of iterations $n_{\max}$.
Randomly initialize $\boldsymbol{\theta}$
$\mathcal{L}_{\mathrm{batch}} \leftarrow +\infty$, $i \leftarrow 0$
while $\mathcal{L}_{\mathrm{batch}} > \epsilon_{\mathrm{tol}}$ do
    Randomly sample a batch $\mathcal{Q}$ of size $B_Q$ from $\{1,2,\ldots,Q\}$
    Compute $\mathcal{L}_{\mathrm{batch}}(\boldsymbol{\theta}) = \sum_{q\in\mathcal{Q}} \left( \left\| |\widetilde{\mathbf{x}}_q| - |\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)| \right\|_1 + \left\| \arg\left(\widetilde{\mathbf{x}}_q\right) - \arg\left(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)\right) \right\|_1 \right)$
    Update $\boldsymbol{\theta}$ with gradient $\boldsymbol{\nabla}_{\boldsymbol{\theta}} \mathcal{L}_{\mathrm{batch}}(\boldsymbol{\theta})$
    $i \leftarrow i+1$
    if $i > n_{\max}$ then
        Exit the while loop
    end if
end while
Output: Network parameters $\boldsymbol{\theta}$
```

3.2. Optimization strategy

The relation between the measurements and the underlying images is nonlinear, which makes the inverse problem very challenging. The fact that only one LED is switched on for each measurement further adds to the difficulty. Thus, in order to avoid bad local minima while solving the optimization problem in (18), we initialize the parameters of the network according to

$$\widetilde{\boldsymbol{\theta}} \in \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^P} \sum_{q=1}^{Q} \left(\left\| |\widetilde{\mathbf{x}}_q| - |\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)| \right\|_1 + \left\| \arg\left(\widetilde{\mathbf{x}}_q\right) - \arg\left(\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)\right) \right\|_1 \right), \tag{19}$$

where $\{\widetilde{\mathbf{x}}_q\}_{q=1}^{Q}$ are low-quality reconstructions obtained via a standard frame-by-frame method. The magnitude $|\cdot|$ and phase $\arg(\cdot)$ operations in (19) are applied component-wise. We can solve (19) using off-the-shelf minibatch stochastic gradient-descent algorithms. However, it is not desirable to run these algorithms till convergence as the network then overfits the artifacts present in the low-quality reconstructions. Thus, in our initialization routine, which is described in algorithm 1, we deploy early stopping by choosing suitable values for the tolerance ϵ_{tol} and the maximum number of iterations $n_{\max}$ (see section 4.2.3 for details).

After the initialization, we can solve (18) using again some minibatch stochastic gradient-descent algorithm. In some cases (for example, when the measurements are corrupted by a non-negligible amount of noise), running the optimization process beyond a certain number of iterations leads to deterioration of the reconstruction quality as the network begins to overfit the measurements. Thus, we also adopt early stopping when necessary.

For both the initialization and reconstruction tasks, we use (minibatch) stochastic gradient-descent algorithms instead of deterministic ones. This introduces additional hyperparameters (batch sizes) that must be set appropriately. However, stochastic methods with small batch sizes require much less memory than the deterministic ones. In fact, if the number of frames Q is large, applying a deterministic gradient-descent method is infeasible. Further, such stochastic methods are also more likely to escape bad local minima and thus reach better solutions. Indeed, in our experiments, we observed that using reasonably small batch sizes ($B_Q = 10$) led to better reconstructions than using large batch sizes ($B_Q = 40$).

Algorithm 2. Joint recovery of dynamic sample and pupil function.

```
Input: Measurements $\{\{\widetilde{\mathbf{y}}_{q,w}\}_{w=1}^{W}\}_{q=1}^{Q}$, LED indices $\{\{l_{q,w}\}_{w=1}^{W}\}_{q=1}^{Q}$, latent vectors $\{\mathbf{z}_q\}_{q=1}^{Q}$, initial
network parameters $\widetilde{\boldsymbol{\theta}}$, initial Zernike coefficients $\widetilde{\mathbf{c}}$, batch sizes $\{B_W, B_Q\}$, number of epochs $n_{\mathrm{ep}}$.
$\boldsymbol{\theta} \leftarrow \widetilde{\boldsymbol{\theta}}$, $\mathbf{c} \leftarrow \widetilde{\mathbf{c}}$
$n_W \leftarrow \lfloor \frac{W}{B_W} \rfloor$, $n_Q \leftarrow \lfloor \frac{Q}{B_Q} \rfloor$
for $n_{\mathrm{ep}}$ epochs do
    for $n_W$ iterations do
        Randomly sample a batch $\mathcal{W}$ of size $B_W$ from $\{1,2,\ldots,W\}$
        for $n_Q$ iterations do
            Randomly sample a batch $\mathcal{Q}$ of size $B_Q$ from $\{1,2,\ldots,Q\}$
            Compute the loss $\mathcal{L}_{\mathrm{batch}}(\boldsymbol{\theta},\mathbf{c}) = \sum_{q\in\mathcal{Q}} \sum_{w\in\mathcal{W}} \mathcal{D}\big(\widetilde{\mathbf{y}}_{q,w}, |\mathbf{H}_{l_{q,w}}(\mathbf{c})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)|^2\big)$
            Update $\boldsymbol{\theta}$ with gradient $\boldsymbol{\nabla}_{\boldsymbol{\theta}}\mathcal{L}_{\mathrm{batch}}(\boldsymbol{\theta},\mathbf{c})$
            Update $\mathbf{c}$ with gradient $\boldsymbol{\nabla}_{\mathbf{c}}\mathcal{L}_{\mathrm{batch}}(\boldsymbol{\theta},\mathbf{c})$
        end for
    end for
end for
Output: Reconstructed images $\{\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)\}_{q=1}^{Q}$, Zernike coefficients $\mathbf{c}$
```

3.3. Recovery of the pupil function

So far, we have assumed complete knowledge of the pupil function in our reconstruction framework. However, the pupil function is typically not well-characterized in FP. Thus, similar to the work in [17, 26], we estimate it along with the sequence of images.

Following [26], we use Zernike polynomials to represent the pupil function with only a few parameters ($\ll M^2$). These functions are orthogonal on the unit circle and are often used in optics for modeling aberrations. We express the pupil function in polar coordinates (ρ, ϕ) as

$$\widehat{p}(\rho,\phi) = \begin{cases} \exp\left(\mathrm{j}\sum_{a=1}^{A} c_a Z_a\left(\frac{\rho\lambda}{2\pi\mathrm{NA}},\phi\right)\right), & \rho \leqslant \frac{2\pi\mathrm{NA}}{\lambda} \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

where Z_a is the ath Zernike polynomial according to Noll's sequential indices (refer to appendix for details) and $\mathbf{c} = (c_a)_{a=1}^{A} \in \mathbb{R}^A$ ($A \ll M^2$) contains the Zernike coefficients. The pupil function is discretized as in (8) by evaluating (20) on the required Cartesian grid. We denote the vectorized discrete pupil function by $\widehat{\mathbf{p}}(\mathbf{c}) \in \mathbb{R}^{M^2}$ to explicitly indicate the dependence on the Zernike coefficients. Similarly, our forward model (9) is then written as

$$\mathbf{y}_{q,w} = |\mathbf{H}_{l_{q,w}}(\mathbf{c})\mathbf{x}_q|^2 = \left|\frac{4\pi^2}{r_{\mathrm{p}}^2}\mathbf{F}_M^{-1}\mathbf{diag}(\widehat{\mathbf{p}}(\mathbf{c}))\mathbf{C}_{\mathbf{k}_{l_{q,w}}}\mathbf{F}_N\mathbf{x}_q\right|^2. \tag{21}$$

Finally, the optimization problem for the joint recovery of the pupil function and the sequence of images is

$$(\boldsymbol{\theta}^*, \mathbf{c}^*) \in \arg \min_{\boldsymbol{\theta}\in\mathbb{R}^P, \mathbf{c}\in\mathbb{R}^A} \sum_{q=1}^{Q}\sum_{w=1}^{W} \mathcal{D}\big(\widetilde{\mathbf{y}}_{q,w}, |\mathbf{H}_{l_{q,w}}(\mathbf{c})\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}_q)|^2\big), \tag{22}$$

where $\mathcal{D} : \mathbb{R}^{M^2} \times \mathbb{R}^{M^2} \to \mathbb{R}_+$ is the data-fidelity term. We can solve (22) using a minibatch stochastic gradient-descent algorithm coupled with early stopping if required. Our complete reconstruction algorithm is summarized in algorithm 2.

4. Numerical results

4.1. Simulated setup

We demonstrate the advantages of our reconstruction method on simulated data. We consider an FP setup consisting of $L = 100$ LEDs arranged in a (10×10) uniform grid with a spacing of $d_L = 4\,\mathrm{mm}$. The maximum illumination NA of the LED array, which is placed at distance $h = 90.88\,\mathrm{mm}$ from the sample, is 0.27. The LEDs emit light with wavelength $\lambda = 532\,\mathrm{nm}$. The NA of the objective is $\mathrm{NA} = 0.1$. We have chosen these values of d_L, h, λ and NA based on the experimental setup in [20]. The pupil function is defined according to (20) using the first nine Zernike polynomials with coefficients $\mathbf{c} = (0, 0.15, 0.3, -0.1, 0.2, 0, 0, 0, 0) \in \mathbb{R}^9$. We take the low-resolution measurements acquired by the camera to be of size (64×64) with pixel-size $\Delta = \frac{\lambda}{4\,\mathrm{NA}} = 1.33\,\mu\mathrm{m}$ and we set the oversampling ratio as $r_\mathrm{p} = 4$. Consequently, the pixel size for the high-resolution image is $\Delta_\mathrm{r} = 332.5\,\mathrm{nm}$ and the step-size for discretizing the pupil function is $\Delta_\mathrm{k} = 0.074\,\mu\mathrm{m}^{-1}$. The LED array and the pupil function are shown in figure 3.

Our ground truth is a sequence of complex-valued images $\{\mathbf{x}_q \in \mathbb{C}^{256^2}\}_{q=1}^{100}$ of size (256×256) which we created from experimental phase images[7]. We place ourselves in the extremely challenging ultrafast regime where only one measurement is acquired for each image in the sequence. For each measurement, a single LED of index[8] l_q is randomly activated and a low-resolution image[8] $\mathbf{y}_q \in \mathbb{R}^{64^2}$ is simulated according to (21). The recorded measurement image $\widetilde{\mathbf{y}}_q \in \mathbb{R}^{64^2}$ is then generated according to

$$\widetilde{\mathbf{y}}_q = \mathbf{y}_q + \mathbf{n}_q, \tag{23}$$

where $\mathbf{n}_q \in \mathbb{R}^{64^2}$ is a realization of a zero-mean Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}_q \in \mathbb{R}^{64^2 \times 64^2}$. Specifically, we consider two settings for our simulations. In the first case, $\boldsymbol{\Sigma}_q$ is the zero matrix, which means that the recorded measurements are noiseless. In the second case, $\boldsymbol{\Sigma}_q$ is a diagonal matrix with entries $\left(([\mathbf{y}_q]_m)/1000\right)_{m=1}^{64^2}$. There, (23) corresponds to a Gaussian approximation of the Poisson noise model with a photon budget of 1000.

We show some of the frames in the ground-truth sequence and the corresponding measurements for both the settings (noiseless and noisy) in figure 4. The full sequences are provided in the supplementary material.

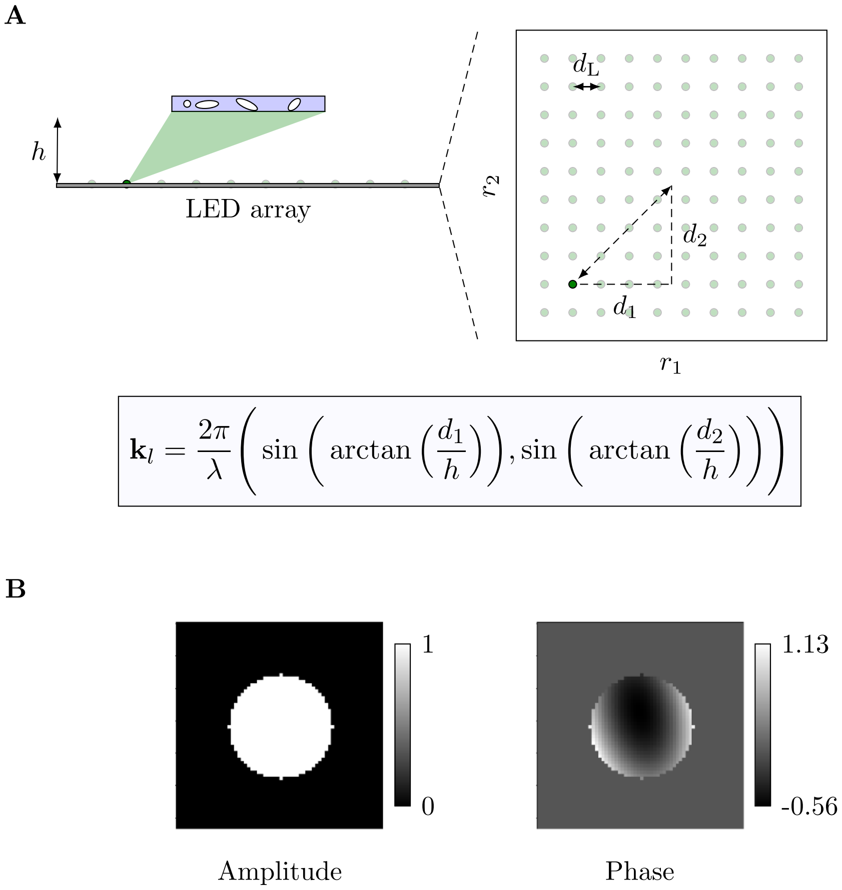
4.2. Implementation of the deep spatiotemporal prior

In this subsection, we describe the implementation of our reconstruction method—the deep spatiotemporal prior (DSTP).

4.2.1. Network architecture. It has been observed that the choice of the network architecture can greatly affect the performance of DIP [40]. Therefore, the common practice when deploying DIP (or other related schemes), is to select the architecture in an empirical trial-and-error manner for the specific task at hand. For our experiments, inspired by [39], we adopt a convolutional decoder-like architecture for $\mathbf{f}_{\boldsymbol{\theta}}$, which, as we demonstrate in sections 4.5 and 4.6,

[7] The experimental phase images are from [41] and are available at http://celltrackingchallenge.net/2d-datasets/.

[8] We have dropped the index w as $W = 1$.



$$\boxed{\mathbf{k}_l = \frac{2\pi}{\lambda}\left(\sin\left(\arctan\left(\frac{d_1}{h}\right)\right), \sin\left(\arctan\left(\frac{d_2}{h}\right)\right)\right)}$$

Figure 3. Simulated FP setup. Panel (A): LED array. Panel (B): Pupil function.

yields high-quality reconstructions. It takes a low-dimensional input vector $\mathbf{z} \in \mathbb{R}^{8^2}$ and outputs a complex-valued (vectorized) image $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{z}) \in \mathbb{C}^{256^2}$. The architectural details are described in table 1. In particular, the complex-valued image is generated from a pair of magnitude and phase images. The initial part of the network creates feature maps of size $(128 \times 256 \times 256)$. These are then fed into both the magnitude and phase branches of the network. The magnitude branch consists of a convolutional layer followed by the pointwise differentiable rectified linear unit (DReLU) activation function, which we define as

$$\mathrm{DReLU}(x) = \begin{cases} \gamma \exp(\frac{x}{\gamma} - 1), & x < \gamma \\ x, & \text{otherwise}, \end{cases} \tag{24}$$

where $\gamma > 0$ is set *a priori*. We use DReLU (with $\gamma = 0.1$) instead of ReLU to avoid the 'dead-neuron' issue during the first few iterations of the optimization, while ensuring that the magnitude is positive. Meanwhile, the phase branch consists of a convolutional layer followed by the π tanh nonlinearity to constrain the phase to lie within the range $[-\pi, \pi]$.

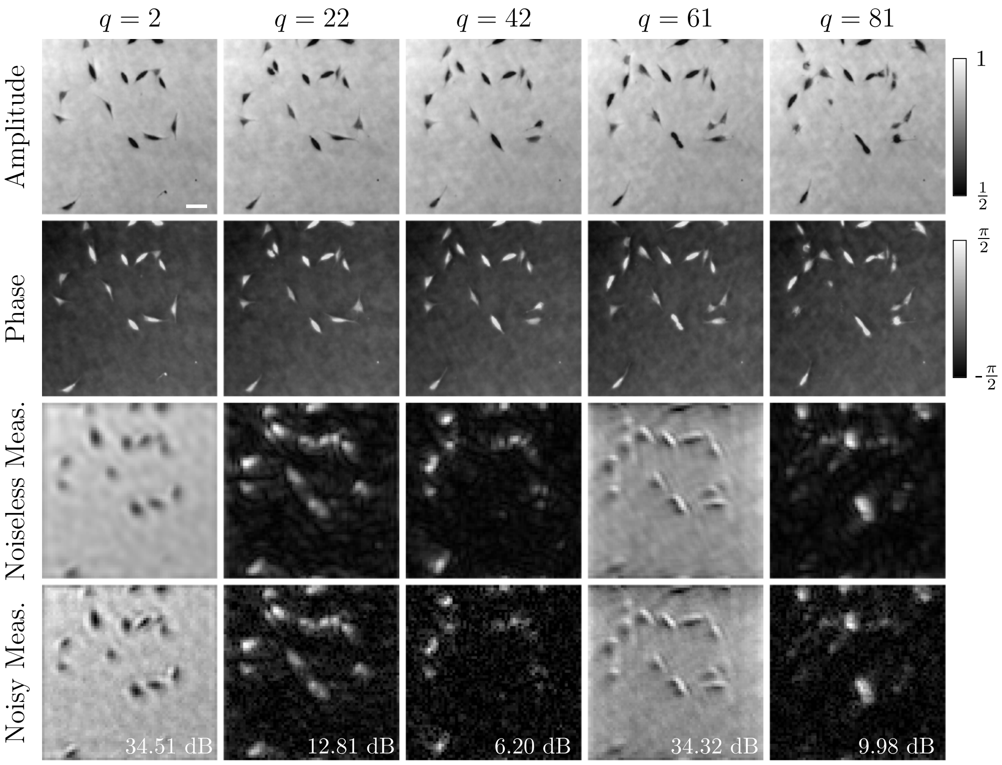


Figure 4. First and second row: frames of the ground-truth sequence (amplitude and phase). Third and fourth row: corresponding low-resolution measurements (noiseless and noisy, normalized for visualization). The signal-to-noise ratios for the noisy measurements, computed as $20\log_{10}\frac{\|\mathbf{y}_q\|_2}{\|\mathbf{y}_q-\tilde{\mathbf{y}}_q\|_2}$, are indicated at the bottom right corners of the measurement images. Scale bar: 10 μm.

4.2.2. Latent vectors. As mentioned in section 3.1, the latent vectors $\{\mathbf{z}_q \in \mathbb{R}^{8^2}\}_{q=1}^{100}$ are chosen such that they lie on the straight line defined in (17). We fix the end-points $\mathbf{z}_1, \mathbf{z}_{100}$ of this line by drawing two samples from the standard multivariate normal distribution in 8^2 dimensions.

4.2.3. Initialization. In all our experiments, we initialize the parameters of the network using reconstructions obtained from the GS algorithm (briefly described in section 4.3). We run algorithm 1 using the AMSGrad solver [42] with a learning rate of 10^{-3}, batch size $B_Q = 10$, tolerance $\epsilon_{\text{tol}} = 0.1 \times (B_Q \times 256^2)$ and maximum number of iterations $n_{\max} = 1000$. We then freeze the tunable parameters of the batch-normalization layers. For experiments involving the estimation of the pupil function, we initialize the Zernike coefficients as $\widetilde{\mathbf{c}} = \mathbf{0}$.

We have observed that the initialization of the network parameters has an impact on the reconstruction quality. For example, randomly initializing the parameters does not lead to satisfactory results. However, initializing the network by simply fitting it to low-quality solutions of the GS algorithm (along with early stopping to avoid overfitting the artifacts) allows us to obtain excellent reconstructions (see sections 4.5 and 4.6).

Table 1. Architecture of the network $\mathbf{f}_{\boldsymbol{\theta}}$. Size of input: (1×8^2). Conv: convolutional layer with (3×3) kernels and reflective boundary conditions. BN: batch normalization layer. Upsampling: nearest neighbor interpolation. The amplitude and phase branches take the same input of size $(128 \times 256 \times 256)$ and output the magnitude and phase images of size $(1 \times 256 \times 256)$, respectively. DReLU is described in (24). The combination layer generates a complex-valued image from the magnitude and phase images. This network consists of 1628 546 learnable parameters.

Layers	Output shape
Reshape	$1 \times 8 \times 8$
$2 \times$ (Conv + BN + ReLU)	$128 \times 8 \times 8$
Upsampling + $2 \times$ (Conv + BN + ReLU)	$128 \times 16 \times 16$
Upsampling + $2 \times$ (Conv + BN + ReLU)	$128 \times 32 \times 32$
Upsampling + $2 \times$ (Conv + BN + ReLU)	$128 \times 64 \times 64$
Upsampling + $2 \times$ (Conv + BN + ReLU)	$128 \times 128 \times 128$
Upsampling + $2 \times$ (Conv + BN + ReLU)	$128 \times 256 \times 256$
Magnitude: Conv + DReLU	$1 \times 256 \times 256$
Phase: Conv + π tanh	$1 \times 256 \times 256$
Combination: Magnitude $\odot \mathrm{e}^{\mathrm{jPhase}}$	$1 \times 256 \times 256$
Reshape	1×256^2

4.2.4. Choice of the data-fidelity term. The data-fidelity term $\mathcal{D}(\cdot,\cdot)$ in (18) and (22) measures the discrepancy between the observed and the simulated measurements. For optimization-based FP reconstruction, it has been experimentally shown in [16] that a cost function of the form

$$\mathcal{D}(\mathbf{a},\mathbf{b}) = \frac{1}{2}\left\|\sqrt{\mathbf{a}} - \sqrt{\mathbf{b}}\right\|_2^2, \tag{25}$$

where $\mathbf{a},\mathbf{b} \in \mathbb{R}^{M^2}$, is robust and leads to better reconstructions than the popular mean-squared-error cost function $\mathcal{D}(\mathbf{a},\mathbf{b}) = \frac{1}{2}\left\|\mathbf{a} - \mathbf{b}\right\|_2^2$. In particular, the cost function in (25) has a gradient similar to that of the Poisson-likelihood-based cost function, which suggests that it can also handle Poisson-like noise well [16]. In our framework, we use the slightly modified version of (25) given by

$$\mathcal{D}(\mathbf{a},\mathbf{b}) = \frac{1}{2}\left\|\sqrt{\mathbf{a}+\epsilon\mathbb{1}} - \sqrt{\mathbf{b}+\epsilon\mathbb{1}}\right\|_2^2, \tag{26}$$

where $\mathbb{1} \in \mathbb{R}^{M^2}$ is a vector with all entries equal to 1 and $\epsilon = 10^{-10}$ helps us avoid numerical instabilities in the computation of the gradient.

Note. The details regarding the optimization process for (18) and (22) are provided in sections 4.5 and 4.6.

4.3. Comparisons

We compare our proposed framework to the following methods.

4.3.1. GS algorithm. The GS algorithm [8] is a classical method for phase retrieval. Assuming that the Zernike coefficient vector $\mathbf{c}$ is known, it aims at solving the feasibility problem

$$\mathbf{x}^*_{\mathrm{GS},q} \in \left\{\mathbf{x} : \tilde{\mathbf{y}}_q = |\mathbf{H}_{l_q}(\mathbf{c})\mathbf{x}|^2\right\} \tag{27}$$

for $q = 1, 2, \ldots, 100$, by alternately updating the image plane and the object plane. We refer the reader to [8] for more details. When the pupil function is not well-characterized, we do not incorporate its recovery within the GS algorithm. Instead, we solve (27) assuming an idealized pupil function with no phase aberrations that corresponds to $\mathbf{c} = \mathbf{0}$.

4.3.2. Data-consistency estimator (DC).

Based on the work in [16], we consider a data-consistency (DC) estimator that minimizes the (slightly modified) 'amplitude-based' cost function (26). For the joint recovery of the images and pupil function, it is given by

$$\left(\mathbf{x}^*_{\mathrm{DC},1}, \ldots, \mathbf{x}^*_{\mathrm{DC},100}, \mathbf{c}^*_{\mathrm{DC}}\right) \in \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_{100},\mathbf{c}} \sum_{q=1}^{100} \mathcal{D}\left(\tilde{\mathbf{y}}_q, |\mathbf{H}_{l_q}(\mathbf{c})\mathbf{x}_q|^2\right), \tag{28}$$

where $\mathcal{D}(\cdot,\cdot)$ is defined in (26).

4.3.3. Spatially total-variation-regularized estimator (STV).

In our numerical simulations, we also consider a regularized estimator where the cost function in (28) is augmented with spatial anisotropic TV regularization for each frame. It is given by

$$\begin{aligned}\left(\mathbf{x}^*_{\mathrm{STV},1}, \ldots, \mathbf{x}^*_{\mathrm{STV},100}, \mathbf{c}^*_{\mathrm{STV}}\right) \in\ & \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_{100},\mathbf{c}} \sum_{q=1}^{100} \Big(\mathcal{D}\left(\tilde{\mathbf{y}}_q, |\mathbf{H}_{l_q}(\mathbf{c})\mathbf{x}_q|^2\right) \\ & + \tau_{\mathrm{amp},q}\left\|\mathbf{L}\{|\mathbf{x}_q|\}\right\|_1 + \tau_{\mathrm{phase},q}\left\|\mathbf{L}\{\arg(\mathbf{x}_q)\}\right\|_1\Big),\end{aligned} \tag{29}$$

where the operator $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{2N}$ computes finite differences in both the directions for the underlying image, and $\{\tau_{\mathrm{amp},q}, \tau_{\mathrm{phase},q}\}_{q=1}^{100} \subset \mathbb{R}_+$ are hyperparameters that control the strength of the regularization.

4.3.4. Spatiotemporally total-variation-regularized estimator (STTV).

Finally, we also implement a spatiotemporally-regularized estimator where the cost function in (28) is augmented with both spatial and temporal TV regularization. It is given by

$$\begin{aligned}\left(\mathbf{x}^*_{\mathrm{STTV},1}, \ldots, \mathbf{x}^*_{\mathrm{STTV},100}, \mathbf{c}^*_{\mathrm{STTV}}\right) \in\ & \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_{100},\mathbf{c}} \sum_{q=1}^{100} \Big(\mathcal{D}\left(\tilde{\mathbf{y}}_q, |\mathbf{H}_{l_q}(\mathbf{c})\mathbf{x}_q|^2\right) \\ & + \tau_{\mathrm{amp},s}\left\|\mathbf{L}\{|\mathbf{x}_q|\}\right\|_1 + \tau_{\mathrm{phase},s}\left\|\mathbf{L}\{\arg(\mathbf{x}_q)\}\right\|_1\Big) \\ & + \sum_{q'=1}^{99} \Big(\tau_{\mathrm{amp},t}\left\||\mathbf{x}_{q'+1}| - |\mathbf{x}_{q'}|\right\|_1 \\ & + \tau_{\mathrm{phase},t}\left\|\arg(\mathbf{x}_{q'+1}) - \arg(\mathbf{x}_{q'})\right\|_1\Big),\end{aligned} \tag{30}$$

where $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{2N}$ is the finite-difference operator and $\{\tau_{\mathrm{amp},s}, \tau_{\mathrm{phase},s}, \tau_{\mathrm{amp},t}, \tau_{\mathrm{phase},t}\} \subset \mathbb{R}_+$ are the regularization hyperparameters.

Table 2. Reconstruction from noiseless measurements with a perfectly characterized pupil function.

Method	GS	DC	STV	STTV	DSTP
RSNR (dB)	17.24	9.66	17.85	18.58	**28.61**

Note: The bold values indicate the method with the best performance.

4.4. Evaluation metric

We quantify the performance of a method by computing the regressed signal-to-noise ratio (RSNR) for the entire reconstructed sequence of images. Let $\overline{\mathbf{x}}$ and $\overline{\mathbf{x}}^*$ denote vectorized versions of the ground truth and reconstruction, respectively. These are created by concatenating the vectorized representations of each frame in the sequence. The RSNR is computed as

$$\mathrm{RSNR}(\overline{\mathbf{x}}^*,\overline{\mathbf{x}}) = \max_{a\in\mathbb{C}} 20\log_{10}\frac{\|\overline{\mathbf{x}}\|_2}{\|\overline{\mathbf{x}}-a\overline{\mathbf{x}}^*\|_2}. \tag{31}$$

We also report the SNR for the pupil function whenever it is jointly estimated with the sequence of images. This metric is computed as

$$\mathrm{SNR}\big(\widehat{\mathbf{p}}(\mathbf{c}),\widehat{\mathbf{p}}(\mathbf{c}^*)\big) = 20\log_{10}\frac{\|\widehat{\mathbf{p}}(\mathbf{c})\|_2}{\|\widehat{\mathbf{p}}(\mathbf{c})-\widehat{\mathbf{p}}(\mathbf{c}^*)\|_2}, \tag{32}$$

where $\mathbf{c}$ and $\mathbf{c}^*$ are the ground-truth and estimated Zernike coefficients, respectively.

4.5. Reconstruction from noiseless measurements

We now present two experiments involving noiseless measurements. In both of them, we run the iterative algorithm for each method for a sufficient number of iterations (details are provided below), beyond which the reconstruction does not change significantly. In other words, we do not deploy early stopping for any method as the measurements are noiseless.

4.5.1. Perfectly characterized pupil function. We first consider an idealized setting where the pupil function is perfectly characterized and is therefore not estimated during the reconstruction of the images of interest. In this scenario, the DC and STV estimators can be computed in frame-by-frame manner (similar to the GS method) by decomposing the overall optimization problems into $Q = 100$ smaller ones. We solve these by running AMSGrad with a learning rate of 10^{-3} for 1000 iterations. In order to improve their performance, we initialize the GS, DC, and STV methods for the timestamp t_q with the reconstructed images from the previous timestamp t_{q-1}. The GS solution is used for initializing the STTV method. We solve (30) by using AMSGrad for 10 000 epochs with a learning rate of 10^{-3} and a full batch size of 100. The optimal hyperparameters $\{\tau_{\mathrm{amp},q},\tau_{\mathrm{phase},q}\}_{q=1}^{100}$ and $\{\tau_{\mathrm{amp},s},\tau_{\mathrm{phase},s},\tau_{\mathrm{amp},t},\tau_{\mathrm{phase},t}\}$ for the STV and STTV methods, respectively, are chosen via a grid-search. For DSTP, the network parameters are initialized with the help of the GS solution. We then solve (18) by running the AMSGrad optimizer for 10 000 epochs with a learning rate of 5×10^{-5} and a batch size of $B_Q = 10$.

We present the RSNR values for all the methods in table 2. Further, we display some slices of the (2D + time) reconstructions in figure 5. The entire reconstructed sequences can be found in the supplementary material. We observe that the proposed method significantly outperforms the GS, DC, STV and STTV methods. Even though only one measurement is acquired per

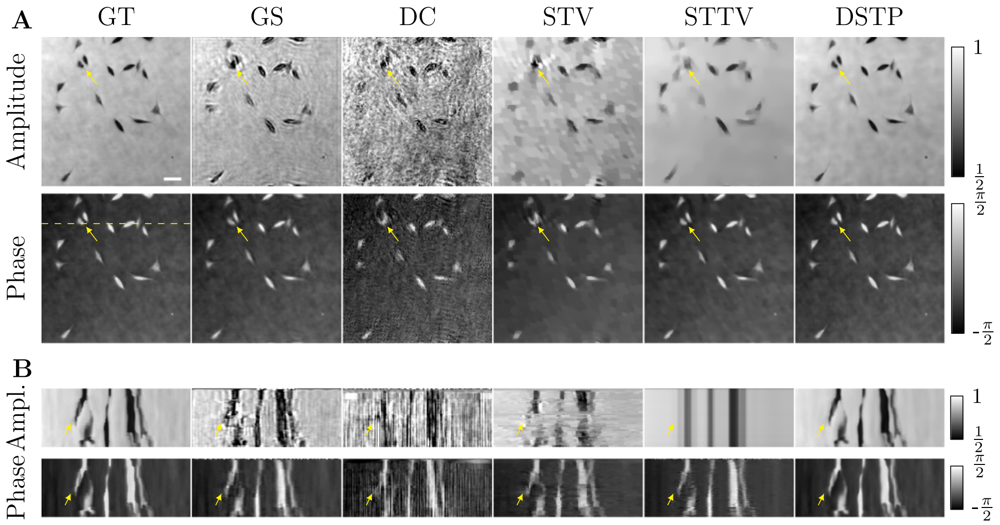


Figure 5. Reconstruction from noiseless measurements with a perfectly characterized pupil function. Panel (A): XY view for the frame index $q = 26$. Panel (B): XT view for the Y position indicated in panel (A) (GT, Phase, dashed line). Scale bar: 10 μm.

Table 3. Joint recovery of the dynamic sample and the pupil function from noiseless measurements.

Method	GS	DC	STV	STTV	DSTP
Sequence RSNR (dB)	14.70	14.82	15.88	17.10	**28.04**
Pupil SNR (dB)	N.A.	8.28	9.57	12.74	**31.22**

Note: The bold values indicate the method with the best performance.

frame, it yields a high-quality reconstruction, unlike the other methods which exhibit various artifacts (for example, the features marked by arrows in figure 5).

4.5.2. Joint recovery of dynamic sample and pupil function. Next, we consider a setting where the pupil function is not well-characterized and is therefore estimated jointly with the dynamic sample in our framework and in the DC, STV and STTV methods. (We do not adapt the GS algorithm for the recovery of the pupil function; we simply assume the idealized pupil function $\mathbf{c} = \mathbf{0}$.) For the DC, STV and STTV methods, the sequence of images is initialized with the GS solution and the Zernike coefficients are initialized as $\widetilde{\mathbf{c}} = \mathbf{0}$. We solve (28) and (29) by running the AMSGrad optimizer for 10 000 epochs with a learning rate of 10^{-3} and a batch size of 10. For solving (30), we run AMSGrad for 10 000 epochs with a learning rate of 10^{-3} and a full batch size of 100. In the STV method, we select two global hyperparameters $\{\tau_{\text{amp}}, \tau_{\text{phase}}\}$ via grid search and share them among all frames. The hyperparameters $\{\tau_{\text{amp},s}, \tau_{\text{phase},s}, \tau_{\text{amp},t}, \tau_{\text{phase},t}\}$ for the STTV method are also tuned for best performance with the help of a grid search. In our method, we initialize the network parameters using the GS solution and we initialize the Zernike coefficients as $\widetilde{\mathbf{c}} = \mathbf{0}$. We solve (18) by running AMSGrad for 10 000 epochs with a learning rate of 5×10^{-5} and a batch size of $B_Q = 10$.

We present the RSNR and SNR values for the reconstructed sequence and the pupil function, respectively, in table 3. We also show some slices of the (2D + time) reconstructions and the recovered pupil functions (phase) in figure 6, as well as the recovered Zernike

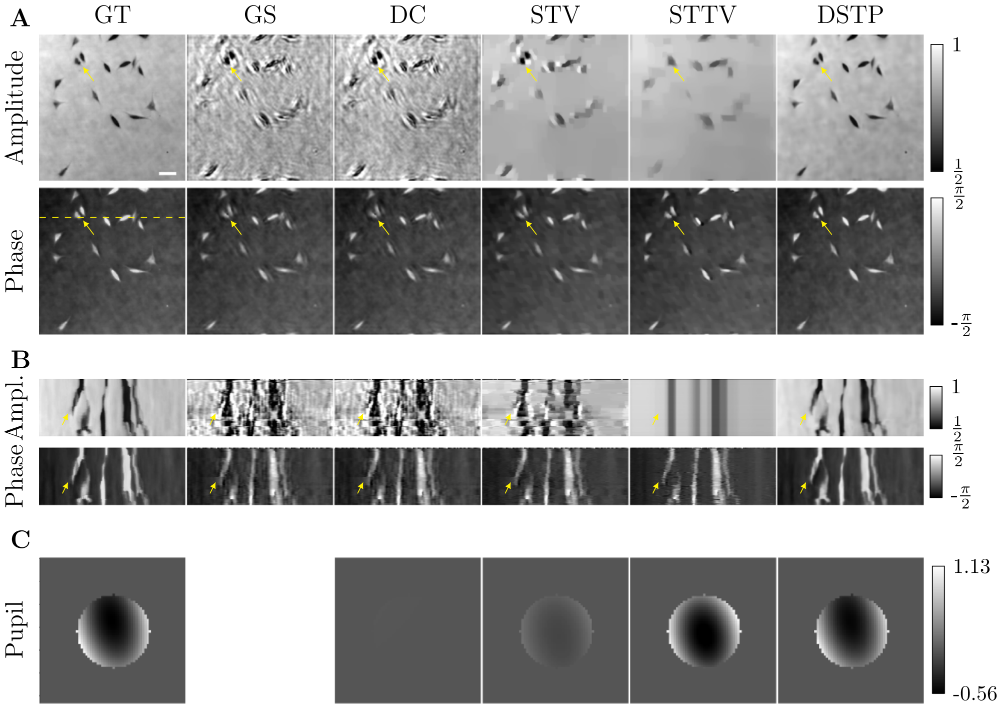


Figure 6. Joint recovery of the dynamic sample and the pupil function from noiseless measurements. Panel (A): XY view for the frame index $q = 26$. Panel (B): XT view for the Y position indicated in panel (A) (GT, Phase, dashed line). Panel (C): phase of the pupil function. Scale bar (for panels (A) and (B)): 10 μm.

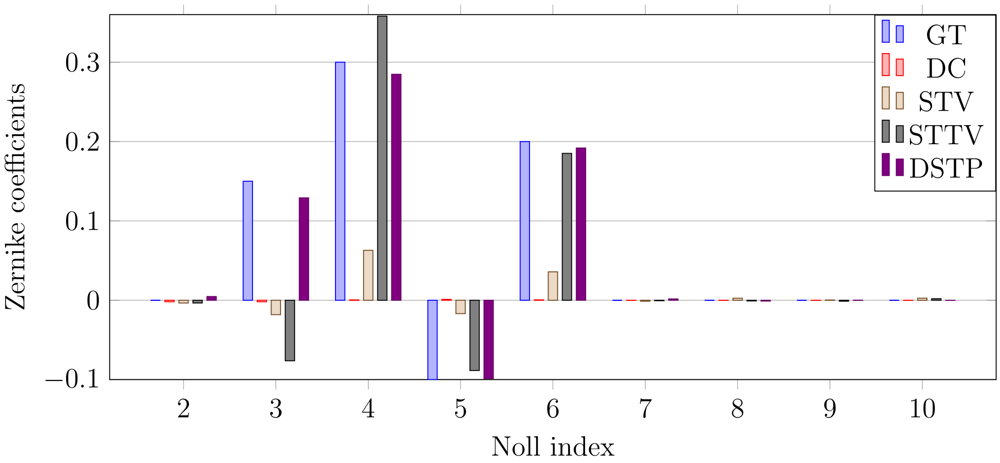


Figure 7. Recovered Zernike coefficients from noiseless measurements. The first (Noll index = 1) Zernike mode only contributes a constant phase factor which has no effect on the intensity measurements and thus can be ignored.

coefficients in figure 7. The full reconstructed sequences are provided in the supplementary material. Here, the DC, STV and STTV methods fail to recover the Zernike coefficients (i.e. the

Table 4. Joint recovery of the dynamic sample and the pupil function from noisy measurements.

Method	GS	DC	STV	STTV	DSTP
Sequence RSNR (dB)	14.09	14.14	14.65	16.39	**24.86**
Pupil SNR (dB)	N.A.	9.36	10.66	14.98	**28.36**

Note: The bold values indicate the method with the best performance.

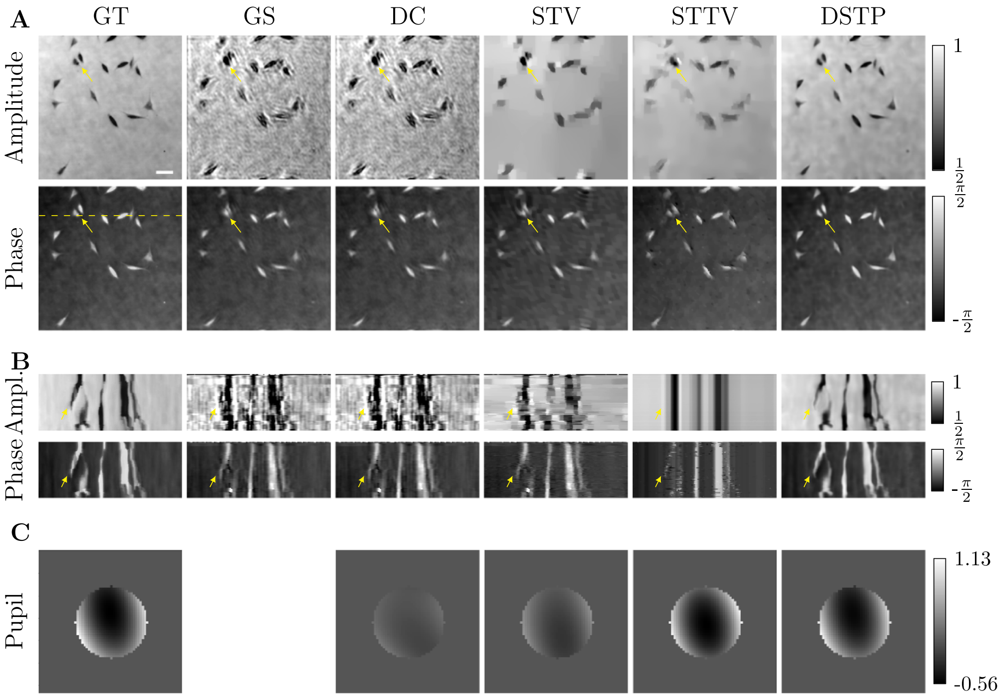


Figure 8. Joint recovery of the dynamic sample and the pupil function from noisy measurements. Panel (A): XY view for the frame index $q = 26$. Panel (B): XT view for the Y position indicated in panel (A) (GT, Phase, dashed line). Panel (C): phase of the pupil function. Scale bar (for panels (A) and (B)): 10 μm.

pupil function) accurately and yield poor reconstructions of the dynamic sample. On the contrary, our method provides a good estimate of the pupil function along with a high-quality reconstruction of the moving sample.

4.6. Reconstruction from noisy measurements

Finally, we consider the joint recovery of the sequence of images and the pupil function from noisy measurements. In this case, we observe that the GS, DC and DSTP methods require early stopping as running the corresponding iterative algorithm beyond a point leads to overfitting the noisy measurements. Thus, we run each method for a sufficiently large number of epochs (=10 000) and we report the reconstruction that achieves the best RSNR during these epochs. For each method, we use the initialization, optimizer, learning rate and batch size described

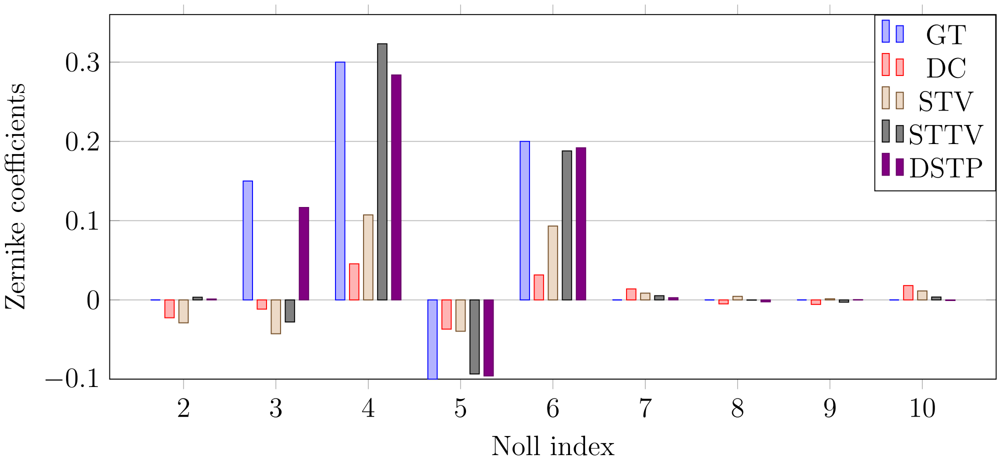


Figure 9. Recovered Zernike coefficients from noisy measurements. The first (Noll index = 1) Zernike mode only contributes a constant phase factor which has no effect on the intensity measurements and thus can be ignored.

in section 4.5.2. The hyperparameters for the STV and STTV methods are also tuned in the same way as in section 4.5.2.

We summarize the quantitative results for all the methods in table 4. We display some slices of the (2D + time) reconstructions and the estimated pupil functions (phase) in figure 8, and we present the recovered Zernike coefficients in figure 9. The entire reconstructed sequences are available in the supplementary material. In this setting, as shown in figure 4, the dark-field measurements are corrupted by significant amounts of noise, which makes the recovery problem quite challenging. Remarkably, our method still yields reconstructions of very good quality and outperforms the DC, STV and STTV methods by a big margin.

4.7. Computational cost

In all our experiments, we used an Intel Xeon Gold 6240R (2.6 GHz) CPU for the GS method and an NVIDIA V100 GPU for the DC, STV, SSTV and DSTP methods. While DSTP achieves substantially better reconstruction quality than the other methods, its computational cost is also higher. For example, the run time for DSTP was around 5.5 h as opposed to $3-30$ min for the other approaches when jointly estimating the sequence and the pupil function from noiseless measurements.

5. Conclusion

We have presented a neural-network-based framework that does not require training data for the reconstruction of high-resolution complex-valued images of a moving sample in dynamic FP. In our method, we have parameterized the sequence of images to be reconstructed using a shared convolutional network with adjustable parameters. We have encoded the temporal behavior of the sample in the input vectors of the network by constraining them to lie on a one-dimensional manifold. In this manner, we have leveraged both the structural prior of a neural network and the temporal regularity between consecutive frames. Further, we have incorporated the recovery of the pupil function of the microscope within our framework. Finally, with the help of simulations, we have shown that the proposed approach drastically improves the quality of reconstruction over standard frame-by-frame methods.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/pakshal23/DynamicFP.

Acknowledgments

This research was supported in part by the Swiss National Science Foundation under Grant 200020_184646/1 and in part by the European Research Council (ERC Project FunLearn) under Grant 101020573.

Appendix. Zernike polynomials

In the polar coordinates (ρ, ϕ), the Zernike polynomials are given by

$$Z_v^u(\rho,\phi) = \begin{cases} R_v^{|u|}(\rho)\cos(|u|\phi), & u \geqslant 0 \\ R_v^{|u|}(\rho)\sin(|u|\phi), & u < 0, \end{cases} \tag{A.1}$$

where $u \in \mathbb{Z}$, $v \in \mathbb{N}$, $\rho \in [0,1]$, $\phi \in [0,2\pi)$, and

$$R_v^{|u|}(\rho) = \begin{cases} \sum\limits_{s=0}^{\frac{(v-|u|)}{2}} \frac{(-1)^s\ (v-s)!}{s!\ \left(\frac{(v+|u|)}{2}-s\right)!\ \left(\frac{(v-|u|)}{2}-s\right)!}\ \rho^{v-2s}, & (v-|u|) \text{ is even,} \\ 0, & (v-|u|) \text{ is odd.} \end{cases} \tag{A.2}$$

For $a \in \mathbb{Z}_+ \setminus \{0\}$, Noll's sequential indexing defines a mapping $Z_v^u \mapsto Z_a$ such that

$$a = \frac{v(v+1)}{2} + |u| + \begin{cases} 0, & u > 0 \ \wedge \ \lfloor v/2 \rfloor \in 2\mathbb{N} \\ 0, & u < 0 \ \wedge \ \lfloor v/2 \rfloor \in 2\mathbb{N}+1 \\ 0, & u \geqslant 0 \ \wedge \ \lfloor v/2 \rfloor \in 2\mathbb{N}+1 \\ 0, & u \leqslant 0 \ \wedge \ \lfloor v/2 \rfloor \in 2\mathbb{N}. \end{cases} \tag{A.3}$$

ORCID iDs

Pakshal Bohra https://orcid.org/0000-0002-2611-3834
Thanh-an Pham https://orcid.org/0000-0001-6231-2569

References

[1] Zheng G, Horstmeyer R and Yang C 2013 Wide-field, high-resolution Fourier ptychographic microscopy *Nat. Photon.* **7** 739–45

[2] Zhang Y, Jiang W, Tian L, Waller L and Dai Q 2015 Self-learning based Fourier ptychographic microscopy *Opt. Express* **23** 18471–86

[3] Fei Cheng Y, Strachan M, Weiss Z, Deb M, Carone D and Ganapati V 2019 Illumination pattern design with deep learning for single-shot Fourier ptychographic microscopy *Opt. Express* **27** 644–56

[4] Tian L, Li X, Ramchandran K and Waller L 2014 Multiplexed coded illumination for Fourier ptychography with an LED array microscope *Biomed. Opt. Express* **5** 2376–89
[5] Tian L, Liu Z, Yeh Li-H, Chen M, Zhong J and Waller L 2015 Computational illumination for high-speed *in vitro* Fourier ptychographic microscopy *Optica* **2** 904–11
[6] Sun J, Zuo C, Zhang J, Fan Y and Chen Q 2018 High-speed Fourier ptychographic microscopy based on programmable annular illuminations *Sci. Rep.* **8** 1–12
[7] Kellman M R, Bostan E, Repina N A and Waller L 2019 Physics-based learned design: optimized coded-illumination for quantitative phase imaging *IEEE Trans. Comput. Imaging* **5** 344–53
[8] Gerchberg R W 1972 A practical algorithm for the determination of phase from image and diffraction plane pictures *Optik* **35** 237–46
[9] Netrapalli P, Jain P and Sanghavi S 2015 Phase retrieval using alternating minimization *IEEE Trans. Signal Process.* **63** 4814–26
[10] Chai A, Moscoso M and Papanicolaou G 2010 Array imaging using intensity-only measurements *Inverse Problems* **27** 015005
[11] Candès E J, Strohmer T and Voroninski V 2013 Phaselift: exact and stable signal recovery from magnitude measurements via convex programming *Commun. Pure Appl. Math.* **66** 1241–74
[12] Waldspurger I, d'Aspremont A and Mallat S 2015 Phase recovery, MaxCut and complex semidefinite programming *Math. Program.* **149** 47–81
[13] Soulez F, Thiébaut É, Schutz A, Ferrari A, Courbin F and Unser M 2016 Proximity operators for phase retrieval *Appl. Opt.* **55** 7412–21
[14] Bian L, Suo J, Chung J, Ou X, Yang C, Chen F and Dai Q 2016 Fourier ptychographic reconstruction using poisson maximum likelihood and truncated Wirtinger gradient *Sci. Rep.* **6** 1–10
[15] Huang Y, Chan A C S, Pan A and Yang C 2019 Memory-efficient, global phase-retrieval of Fourier ptychography with alternating direction method *Computational Optical Sensing and Imaging (COSI 2019)* p CTu4C–2
[16] Yeh L-H, Dong J, Zhong J, Tian L, Chen M, Tang G, Soltanolkotabi M and Waller L 2015 Experimental robustness of Fourier ptychography phase retrieval algorithms *Opt. Express* **23** 33214–40
[17] Ou X, Zheng G and Yang C 2014 Embedded pupil function recovery for Fourier ptychographic microscopy *Opt. Express* **22** 4960–72
[18] Sun J, Chen Q, Zhang Y and Zuo C 2016 Efficient positional misalignment correction method for Fourier ptychographic microscopy *Biomed. Opt. Express* **7** 1336–50
[19] Eckert R, Phillips Z F and Waller L 2018 Efficient illumination angle self-calibration in Fourier ptychography *Appl. Opt.* **57** 5434–42
[20] Zheng G, Shen C, Jiang S, Song P and Yang C 2021 Concept, implementations and applications of Fourier ptychography *Nat. Rev. Phys.* **3** 207–23
[21] Kuang C, Ma Y, Zhou R, Lee J, Barbastathis G, Dasari R R, Yaqoob Z and So P T C 2015 Digital micromirror device-based laser-illumination Fourier ptychographic microscopy *Opt. Express* **23** 26999–7010
[22] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[23] Ren D, Bostan E, Yeh Li-H and Waller L 2017 Total-variation regularized Fourier ptychographic microscopy with multiplexed coded illumination *Imaging and Applied Optics 2017* p MM3C.5
[24] Shi Q, Hui W, Huang K, Zhao H, Ye Q, Tian J and Zhou W 2021 Under-sampling reconstruction with total variational optimization for Fourier ptychographic microscopy *Opt. Commun.* **492** 126986
[25] Zhang Y, Liu Y, Jiang S, Dixit K, Song P, Zhang X, Ji X and Li X 2021 Neural network model assisted Fourier ptychography with Zernike aberration recovery and total variation constraint *J. Biomed. Opt.* **26** 1–14
[26] Zhang Y, Liu Y, Li X, Jiang S, Dixit K, Zhang X and Ji X 2019 PgNN: physics-guided neural network for Fourier ptychographic microscopy (arXiv:1909.08869)
[27] Jagatap G, Chen Z, Hegde C and Vaswani N 2018 Sub-diffraction imaging using Fourier ptychography and structured sparsity *2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 6493–7
[28] Sun Y, Xu S, Li Y, Tian L, Wohlberg B and Kamilov U S 2019 Regularized Fourier ptychography using an online plug-and-play algorithm *2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 7665–9
[29] Dabov K, Foi A, Katkovnik V and Egiazarian K 2007 Image denoising by sparse 3-D transform-domain collaborative filtering *IEEE Trans. Image Process.* **16** 2080–95

[30] McCann M T, Hwan Jin K and Unser M 2017 Convolutional neural networks for inverse problems in imaging: a review *IEEE Signal Process. Mag.* **34** 85–95

[31] Ongie G, Jalal A, Metzler C A, Baraniuk R G, Dimakis A G and Willett R 2020 Deep learning techniques for inverse problems in imaging *IEEE J. Sel. Areas Inf. Theory* **1** 39–56

[32] Kappeler A, Ghosh S, Holloway J, Cossairt O and Katsaggelos A 2017 Ptychnet: CNN based Fourier ptychography *2017 IEEE Int. Conf. on Image Processing (ICIP)* pp 1712–6

[33] Nguyen T, Xue Y, Li Y, Tian L and Nehmetallah G 2018 Deep learning approach to Fourier ptychographic microscopy *Opt. Express* **26** 26470–84

[34] Zhang J, Xu T, Shen Z, Qiao Y and Zhang Y 2019 Fourier ptychographic microscopy reconstruction with multiscale deep residual network *Opt. Express* **27** 8612–25

[35] Shamshad F, Abbas F and Ahmed A 2019 Deep ptych: subsampled Fourier ptychography using generative priors *2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 7720–4

[36] Shamshad F, Hanif A, Abbas F, Awais M and Ahmed A 2019 Adaptive ptych: leveraging image adaptive generative priors for subsampled Fourier ptychography *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV) Workshops*

[37] Konda P C, Loetgering L, Zhou K C, Xu S, Harvey A R and Horstmeyer R 2020 Fourier ptychography: current applications and future promises *Opt. Express* **28** 9603–30

[38] Chen Z, Jagatap G, Nayer S, Hegde C and Vaswani N 2018 Low rank Fourier ptychography *2018 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 6538–42

[39] Yoo J, Jin K H, Gupta H, Yerly J, Stuber M and Unser M 2021 Time-dependent deep image prior for dynamic MRI *IEEE Trans. Med. Imaging* **40** 3337–48

[40] Ulyanov D, Vedaldi A and Lempitsky V 2018 Deep image prior *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 9446–54

[41] Rapoport D H, Becker T, Mamlouk A M, Schicktanz S and Kruse C 2011 A novel validation algorithm allows for automated cell tracking and the extraction of biologically meaningful parameters *PLoS One* **6** 1–16

[42] Reddi S J, Kale S and Kumar S 2018 On the convergence of adam and beyond *Int. Conf. on Learning Representations*