



OPEN

Holographic optical field recovery using a regularized untrained deep decoder network

Farhad Niknam^{1✉}, Hamed Qazvini¹ & Hamid Latifi²

Image reconstruction using minimal measured information has been a long-standing open problem in many computational imaging approaches, in particular in-line holography. Many solutions are devised based on compressive sensing (CS) techniques with handcrafted image priors or supervised deep neural networks (DNN). However, the limited performance of CS methods due to lack of information about the image priors and the requirement of an enormous amount of per-sample-type training resources for DNNs has posed new challenges over the primary problem. In this study, we propose a single-shot lensless in-line holographic reconstruction method using an untrained deep neural network which is incorporated with a physical image formation algorithm. We demonstrate that by modifying a deep decoder network with simple regularizers, a Gabor hologram can be inversely reconstructed via a minimization process that is constrained by a deep image prior. The outcoming model allows to accurately recover the phase and amplitude images without any training dataset, excess measurements, or specific assumptions about the object's or the measurement's characteristics.

Approach overview and background. Digital holographic microscopy is one of the most widely explored modalities as it permits high-throughput multi-dimensional imaging of phase and amplitude information of the specimen. In the meantime, Gabor-based Lensless In-line Holographic Microscopy (LIHM)^{1–6} has attracted special attention because of its simplicity, compactness, and high space-bandwidth product. In any coherent holographic configuration, it is typical to have coherence-related noises such as speckle, defocus, self/cross-interference artifacts. However, due to incomplete data acquisition by digital image sensors which only contain the intensity information of the complex optical field, the LIHM reconstruction gives rise to images overlaid by a spatial artifact called *twin image*. Although this problem can be robustly solved in off-axis geometry by angled illumination which involves an optical setup^{7–10}, this effect can be computationally eliminated by imposing physical constraints that the twin image does not satisfy. Such constraints can be performed through iterative error reduction procedures known as phase retrieval^{11–13}.

In any phase retrieval layout, information diversity must be conducted. Then, the problem can be solved using the excess information as a prior which constrains the possible solutions by reducing the number of unknowns. For instance, one can record multiple holograms in different sample-to-sensor distances and recover the complete object information through a physics-based iterative process^{6,14,15}. Such multi-holograms can also be retrieved from multiple wavelengths¹⁶, angles¹⁷, or phase shifts^{18,19}; generally derived from the methods of alternating projections^{11–13,20,21}. Despite their robustness, *Multiple measurements* are the backbone of these approaches that limits the usage of LIHM for different imaging problems, especially with poor image acquisition rate. Furthermore, these techniques are susceptible to instrumentation errors or other undesired sample-specific or per measurement of environmental noises.

Another strategy is to inversely estimate the complex-valued object field. By integrating known prior information about the object in the reconstruction process such as object-support²² or bound and sparsity constraints^{23–26} in the object's original or transformed domain, the desired optical properties of the object could be retrievable. However, since such an approach requires handcrafted image priors and perfectly tuned parameters, this strategy is limited to very specific small, simple, or undetailed categories of objects.

In recent years, deep neural networks (DNN) trained in an end-to-end fashion on large datasets have been used for phase retrieval, directly mapping measured intensities to fully resolved object fields. Different studies have demonstrated the state-of-the-art performance of DNNs in various imaging problems^{27–31} such as LIHM^{32,33}.

¹Laser and Plasma Research Institute, Shahid Beheshti University, Tehran 1983963113, Iran. ²Department of Physics, Shahid Beheshti University, Tehran 1983963113, Iran. ✉email: farhad.niknam.em@gmail.com

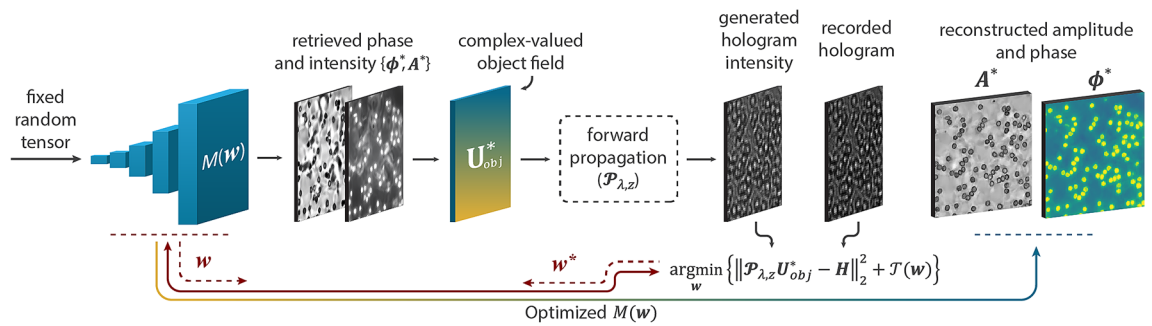


Figure 1. An overview of the proposed reconstruction algorithm. The inverse problem can be solved by optimizing a network of weights by the mean-squared loss between the intensity of the generated hologram and the recorded image in an iterative gradient descent procedure.

In such techniques, a neural network learns the correspondence between the measured and the ground-truth information through a training process. Then, based on the trained image-to-image mapping features, the network could map the different input measured images to approximations of their ground-truth equivalents. However, for nonconvex nonlinear problems, DNNs would learn specific characteristics of the training data that do not necessarily map to the desired distribution for the experimental data. This is a known issue in all data-driven reconstruction strategies that may lead to failures when predicting patterns that were not provided in the training dataset³⁴.

Here we demonstrate a new reconstruction algorithm for LIHM based on a generative untrained deep neural network which is incorporated with a physical model. Inspired by recent successful implementations on several phase retrieval problems^{35,36}, this approach is based on the idea of image generation using a randomly initialized DNN; the so-called deep image prior (DIP) devised by Ulyanov et al.³⁷. Based on this concept, if a randomly initialized deep network is optimized to fit an image, the convolution layers would act as a natural image prior which enables the network to restore the image from defects such as noises and artifacts.

Using the concept of DIP, in Ref.³⁵ the authors demonstrated that an off-axis hologram could be reconstructed by fitting an untrained encoder–decoder network to the hologram through the holographic reconstruction algorithm. Though their implementation is only valid when the absorbance is negligible, and the object field can be solely described in terms of phase. In Ref.³⁶, it is demonstrated that not only the DIP can be incorporated with a physical model to recover the phase information of an object, but also it expands the possibilities of involving other complicated constraints e.g., solving problems that are deeply connected to the image formation process such as optical aberrations.

In the context of lensless in-line holographic reconstruction, the proposed approach in³⁵ is applicable to transparent phase-only samples. Assuming the amplitude distribution is uniform, the indeterminism in the phase recovery process will be minimized. However, for real-life applications, there are limited choices as phase-only substances. Most tissues and living cells exhibit at least some portion of absorption in the visible spectrum regardless of other optical effects like refraction.

Although the acquired information is not enough to fully resolve the object, as explained in³⁶, a DIP model can be versatile enough to accept externally defined priors to overcome the under-sampling problem of in-line holography. These priors could be generic constraints on the network parameters or their outcomes.

For instance, a simple under-parameterized network (has fewer weights than the number of image pixels) is proposed in³⁸ called deep decoder network (DDN) which is also exploited in³⁶. Thanks to its few parameters, this network naturally performs a regularization by simplifying the representation of the image signal. DDNs are simple, robust, and do not require any early stopping. Another advantage is their stronger regularization imposed by their learned priors on the basis of fewer parameters. This makes them more resilient to noise in optimization problems compared to complicated convolution networks such as encoder-decoders. Furthermore, a DDN is basically initialized by a constant random tensor. This feature of fixed input enables more control over the network parameters and its outputs, which allows performing scheduled variation in inputs and parameters without worrying about any side effects such as chaotic behavior or instability.

The under-parameterization of DDNs -alone- is not enough to confront the indeterminism of the problem. However, by introducing a particular set of constraints on the penalty function, it is possible to modify the DIP of the deep network such that the model could effectively recover the object information.

In this research, we demonstrate that by utilizing simple and commonly available regularization methods, a DDN with a physical image formation algorithm can be upgraded to a powerful compressive signal reconstruction model. The proposed model can robustly recover amplitude and phase information of Gabor holograms with state-of-the-art performance using far fewer samples than required by the Nyquist criterion. For abbreviation, we name our approach Deep Compressed Object Decoder or DCOD.

Reconstruction theory and principles of image formation. In the proposed algorithm (Fig. 1), a randomly initialized DDN (as a universal model) predicts the phase and amplitude images. After constructing a hologram by forward propagation of the estimated object field, the weights of the DDN are updated based on the error between the generated and the recorded holograms in addition to a regularization term. Here, the

principles of image formation and the corresponding inverse problem will be described. Following this first step, we will provide a detailed account of the practical implementation.

In general, a thin transparent object at $z = 0$ plane can be characterized by a planar complex-valued object function as:

$$U_{obj}(x, y; z = 0) = A(x, y) \exp(j\phi(x, y)) \quad (1)$$

where A and ϕ represent the transmittance and phase responses of the object. By considering an incident coherent wavefront $U_{inc}(x, y; z = 0)$ on the object plane, with a wavelength of λ , at the sensor plane which is placed at a distance of z_s away from the object, the transmitted and propagated field U can be described using the angular spectrum of the object field on the sensor plane⁷ as:

$$U(x, y; z = z_s) = \mathcal{F}^{-1} \{ P(\mathbf{v}, \lambda, z_s) \cdot \mathcal{F}(U_{inc}(x, y; 0) U_{obj}(x, y; 0)) \} \quad (2)$$

where \mathcal{F}^{-1} and \mathcal{F} are inverse Fourier and Fourier transform operators. $P(\cdot)$ is the free-space optical transfer function that gives a complex-valued matrix and depends on spatial frequency components $\mathbf{v} = (v_x, v_y)$, wavelength λ , and distance of propagation z . $P(\mathbf{v}, \lambda, z)$ can be stated as:

$$P(\mathbf{v}, \lambda, z) = \exp\left(\frac{2\pi j z}{\lambda} \sqrt{1 - (\lambda v_x)^2 - \lambda v_y^2}\right) \quad (3)$$

Now assuming the optical field on the sensor plane only includes a diffraction pattern of the object and by considering $U_{inc}(x, y; 0) U_{obj}(x, y; 0) = U_0$ in Eq. (2), the intensity of the measured image can become:

$$I(x, y; z = 0) = |U(x, y; z = 0)|^2 \equiv \mathbf{P}_{\lambda, z} \mathbf{U}_0 : \mathbf{H} \quad (4)$$

$\mathbf{P}_{\lambda, z}$ is the propagation operator and \mathbf{H} is the so-called hologram^{7,39–41}.

Equation (4) can be solved by inversely finding an estimate of the object distribution \mathbf{U}_{obj}^* for which the corresponding diffraction field \mathbf{U}^* has minimum difference with the ground-truth diffraction field \mathbf{U} . This can be carried out using iterative error reduction algorithms such as gradient descent. The DIP approach states that we can make this approximation with a deep network and optimize it by minimizing a penalty function. The idea behind DIP is that the feature extraction process in a randomly initialized untrained convolution neural network forms an optimizable prior that perfectly fits for each reconstruction. Thus, there exists an optimized set of parameters \mathbf{w} for which the DNN model $M(\mathbf{w})$, can estimate the amplitude A^* and phase ϕ^* distributions of the object \mathbf{U}_{obj}^* whose diffraction intensity at the sensor plane $|\mathbf{U}^*|^2$ has minimum difference with the recorded hologram \mathbf{H} . This can be formally expressed by a minimization problem with mean squared loss as follows:

$$f(\mathbf{w}^*) = \operatorname{argmin}_{\mathbf{w}} \left\{ \|\mathbf{P}_{\lambda, z} \mathbf{U}_{obj}^* - \mathbf{H}\|_2^2 + \mathcal{T}(\mathbf{w}) \right\} \quad (5)$$

$f(\mathbf{w}^*)$ is the minimized loss function that returns the optimal \mathbf{w}^* argument for which $M(\mathbf{w}^*)$ gives the reconstructed images. Assuming $U_{inc} = \mathbf{1}$, then $\mathbf{U}^* = \mathbf{P}_{\lambda, z} \mathbf{U}_{obj}^*$ where \mathbf{U}_{obj}^* can be expressed as Eq. (1), whereas $\phi^* = M_\phi(\mathbf{w})$ and $A^* = M_A(\mathbf{w})$. M_ϕ and M_A are the dual output channels of the network M . $\mathcal{T}(\mathbf{w})$ is an externally defined regularization term. It is clear that the ground-truth object function U_{obj} did not appear in Eq. (5) and the optimization process is only driven by a single measurement, \mathbf{H} . Note that the periodic nature of the complex exponential phase term in Eq. (1), causes the recovered phase to be limited to the range $[-\pi, \pi]$; resolving this limitation is beyond the scope of this study.

For thin and transparent samples, it is possible to encounter the problem as phase retrieval of a phase object. This assumption immediately eliminates the issue of incomplete measurement and allows to solve Eq. (4) without defining any specific regularization. Assuming the amplitude term is mostly uniform and has a small contribution in the object field, it can be approximated as a uniform matrix filled with the average value of the background. Consequently, the model $M(\mathbf{w})$ only gives an estimate of the phase matrix ϕ^* while its corresponding amplitude becomes $A^* = \mathbf{I}$. \mathbf{I} is a constant matrix in which $I_{ij} = s \in \mathbb{R}$ for all $i, j \in \{0, \dots, n_d - 1\}$ where $s = \text{mean}(\text{background})$. It is based on the PhysenNet framework proposed in³⁵ which is implemented on a U-Net, and its minimization does not require the regularization term in Eq. (5).

Resolving the incomplete signal using a regularized DDN. Since the measured signal U is incomplete regardless of the level of noise, Eq. (4) is ill-posed and Eq. (5) becomes hard to solve. The most important issue here is the twin image problem which cannot be suppressed by reducing the spatial details of the object to the features. Such a source of noise imposes a severe level of nonlinearity in Eq. (4) which overlays on the object diffraction field across the entire space and requires carefully defined priors to penalize. Thickness or axial distribution of the object are other issues that add further complication to Eq. (5). Despite the regularization of DDN, without any strongly confining priors on the solutions, the network has the capacity to produce noise. The purpose of the regularization term $\mathcal{T}(\mathbf{w})$ in Eq. (5) is to address this problem. This term may be expanded to the following arguments:

1. Assuming the values of the solution are limited to a particular range, one simple prior would be a bound constraint on the solutions. Such constraint can be easily fulfilled by applying a bounded activation function on the output layer such as sigmoid. This function can limit the resulting values of each amplitude and phase solution to $[0, 1]$.

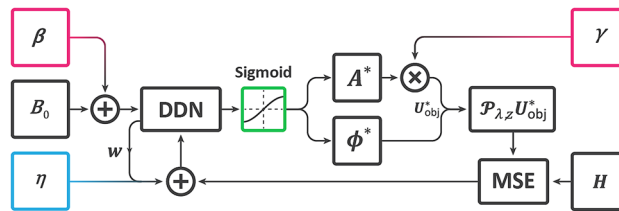


Figure 2. Structure of the DCOD algorithm. Colored boxes are the regularization elements.

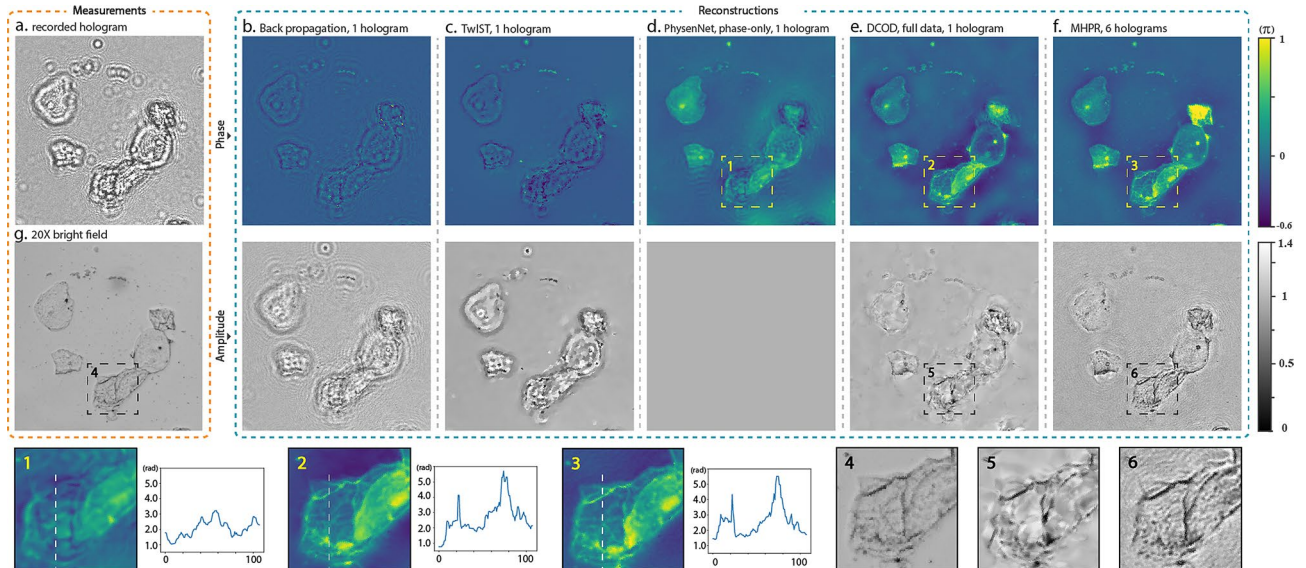


Figure 3. (a) The recorded hologram. (b) The backpropagated phase and amplitude images. (c) The images obtained by TwIST algorithm. (d) The reconstructed phase by PhysenNet assuming the object function only contains phase information. (e) The recovered phase and amplitude images by the proposed DCOD algorithm. (f) The reconstructed images obtained by multi-height phase recovery algorithm. (g) The bright-field microscopic image captured using a 20 \times objective lens (0.4NA).

- Assuming the variables are small, independent, and uncorrelated, energy minimization regularization algorithms such as the ℓ_2 regularization or the so-called weight decay reduces (or removes) the contribution of every non-significant component in the features space which substantially improves the reconstruction outcome.
- Since the loss function is non-convex, minimizing the cost may not necessarily lead to a true solution. Assuming the loss function has a global minimum, perturbing the parameters periodically could prevent overfitting on a particular set of parameters.

About the latter argument, note that our purpose is to fit the model output to a single measured image signal. As a result, overfitting the noise is very probable even with the presence of the powerful DDN image prior and the weight decay in the reconstruction process. Information diversity in supervised neural networks is proved to be an effective way to generalize the trained models and to prevent overfitting. Although it is impossible to define such diversity without any dataset, by applying scheduled random perturbations to the weights, the model tries to optimize the network on a slightly different set of parameters. Random noise could be added directly to the model parameters, to the input tensor of the DDN, or the outputs of the network.

Considering γ as a positive scalar variable, β as a tensor variable with the same shape of B_0 which is the input tensor of the network, and η as the weight decay, the model could be visualized as Fig. 2. γ and β are randomizing variables that could be randomly changed during the optimization to perturb the network parameters.

Results

Figure 3 shows the recorded intensity image (Fig. 3a) and its reconstructed phase and amplitude images of a sample of smeared unstained cheek-cells placed at 238 μm from the sensor plane.

The backpropagation results shown in Fig. 3b are directly solved by propagating the hologram to the object plane, or $U_{rec} = P_{\lambda, z} H$. The reconstructed object information in U_{rec} is disturbed and highly entangled in the overlaying intense twin image noise which is the result of incomplete data acquisition. In summary, Fig. 3b is a visualization of the fundamental ill-posedness of the problem.

Figure 3c depicts the results obtained by TwIST (Two-step Iterative Shrinkage/Thresholding⁴²) algorithm accompanied with a TV-regularizer (Supplementary section A). This approach is widely used to compressively solve inverse problems as holographic reconstruction^{24,25,43}. Due to the small sample-to-sensor distance of the adopted LIHM experimental setup, the field information is encoded within intensely entangled interferometric patterns from which, TwIST algorithm cannot properly reconstruct complex geometries. In fact, the outcomes are merely TV-denoised versions of the backpropagated images.

Figure 3d is obtained by a PhysenNet which is made on top of a DDN as the decoder unit and is initialized with a random tensor. The deep network is followed by the computational propagation of the generated object field creating the output hologram. Assuming the amplitude of the object is uniform, the network-made phase image is combined with a uniform matrix as the amplitude to form the complex-valued object function [Eq. (1)]. The problem is then solved using the Adam optimization algorithm⁴⁴, which minimizes the Mean Squared Error (MSE) of the generated holograms. Finally, optimization is performed with a learning rate of 0.01 for 20,000 iterations. The phase-only constraint, although simple and effective, allows little flexibility against any model mismatch imposed by amplitude variations and thus, fails to resolve the object phase.

Figure 3e illustrates the reconstructed images obtained by the proposed algorithm and the object field is fully recovered using only a single hologram. The outcomes are comparable with the images obtained by the Multi-Height Phase Recovery method (MHPR, Fig. 3f) and imply that the proposed model outperforms TwIST. The optimization of the DCOD is carried out using a weight decay-included Adam optimizer called AdamW⁴⁵ with a learning rate of 0.01 and a weight decay of 0.002 for 35,000 iterations. The results obtained by this algorithm have excellent agreement with the images acquired by MHPR (using 6 axially spaced holograms) and bright field microscopy (using a 20× objective lens, Fig. 3g). The regularization steps taken to overcome the nonlinearity are discussed thoroughly in the subsequent section.

It is worth mentioning that the model clearly failed to recover the bright spot in Fig. 3f that has a large phase value ($> \pi$). In such circumstances, the optimizer gets confused since we did not define any mechanism to perform phase unwrapping.

Discussion

Regularization techniques. Since the inverse problem in in-line holography is inherently nonconvex and nonlinear, besides applying positivity and bound constraints on the solutions, it is crucial to penalize complexity by performing further regularizations to help the model converge.

We applied different techniques to improve convergence. One common way is to contribute the weights in the loss function multiplied by a coefficient called weight decay⁴⁶ to encourage the weights to become small. For Stochastic Gradient Decent (SGD) algorithm, weight decay regularization is equivalent to ℓ_2 regularization of the loss function. This can be formally expressed regarding $\frac{\eta}{2\alpha} \|\mathbf{w}\|_2^2$ as the regularization term $\mathcal{T}(\mathbf{w})$ in Eq. (5), where η and α are weight decay and learning rate parameters respectively. As stated in⁴⁵, For adaptive gradient-based methods (Adam⁴⁴, AdaGrad⁴⁷, AMSGrad⁴⁸, etc.), this resemblance is not valid since with ℓ_2 regularization, the regularizer takes the sums of the gradient of the loss function and the gradient of the regularizer into account, while with weight decay regularization, only the gradient of the loss function needs to be adopted with the regularizer. Therefore, a modified variant of Adam optimizer called AdamW is proposed that decouples loss-based gradient updates in Adam and weight decay, which substantially improves the regularization efficiency⁴⁵. It is demonstrated that decoupled weight decay regularization not only provides flexibility for hyperparameter tuning, but also improves generalization for different experimental settings.

In order to further improve convergence, regularization is extended by introducing randomness to the model parameters. We applied randomization in two ways: first, by adding random noise to the initial fixed random tensor, and second, by multiplying a periodically varying coefficient (> 1) to the generated intensity image, which effectively perturbs the parameters of the network. Periodic random perturbations enable the neural network to give more emphasis to robust features across detailed environments. We observed that applying random perturbations to the model parameters significantly accelerates convergence and gives rise to reconstructions with more meaningful information.

To investigate the influence of these modifications, another experiment is conducted on smeared red blood cells; the results of which are shown in Fig. 4. As depicted in Fig. 4a, the model failed to find the solution with unsophisticated implementation of DDN with Adam optimizer. In Fig. 4b as well, due to lack of randomization, the true solution is lost across a broad range of possible solutions giving similar diffraction intensities on the hologram plane. Unlike the previous cases, the images generated by our algorithm in Fig. 4c,d have clear equivalence with the images obtained by the MHPR method (Fig. 4e). Specifically, the images in Fig. 4c are general representations of phase and amplitude that are regularized by randomization which was applied by perturbing the network every 500 iterations for 30,000 iterations. The initial output is over smoothed, hence with 5000 more iterations using a lower weight decay (0.1 of the initial value) and without randomization, more details will appear (Fig. 4d).

Figure 4f shows MSE variations during the optimization for different regularization settings mentioned. The spike that appeared on about 30,000 iterations on the AdamW graph (orange curve) is evidence of instability. Due to the lack of scheduled reduction of weight decay and learning rate during training, AdamW becomes unstable after several thousand iterations and requires early stopping. This instability is more obvious after letting the optimization run up to 100,000 iterations (see Supplementary Fig. S1). Another benefit of randomization is to stabilize the optimization when weight decay regularization is applied.

The generated holograms at the sensor plane shown in Fig. 4 are hardly distinguishable. This is the result of overfitting when regularization is not applied appropriately. Furthermore, the PSNR and MSE values are not

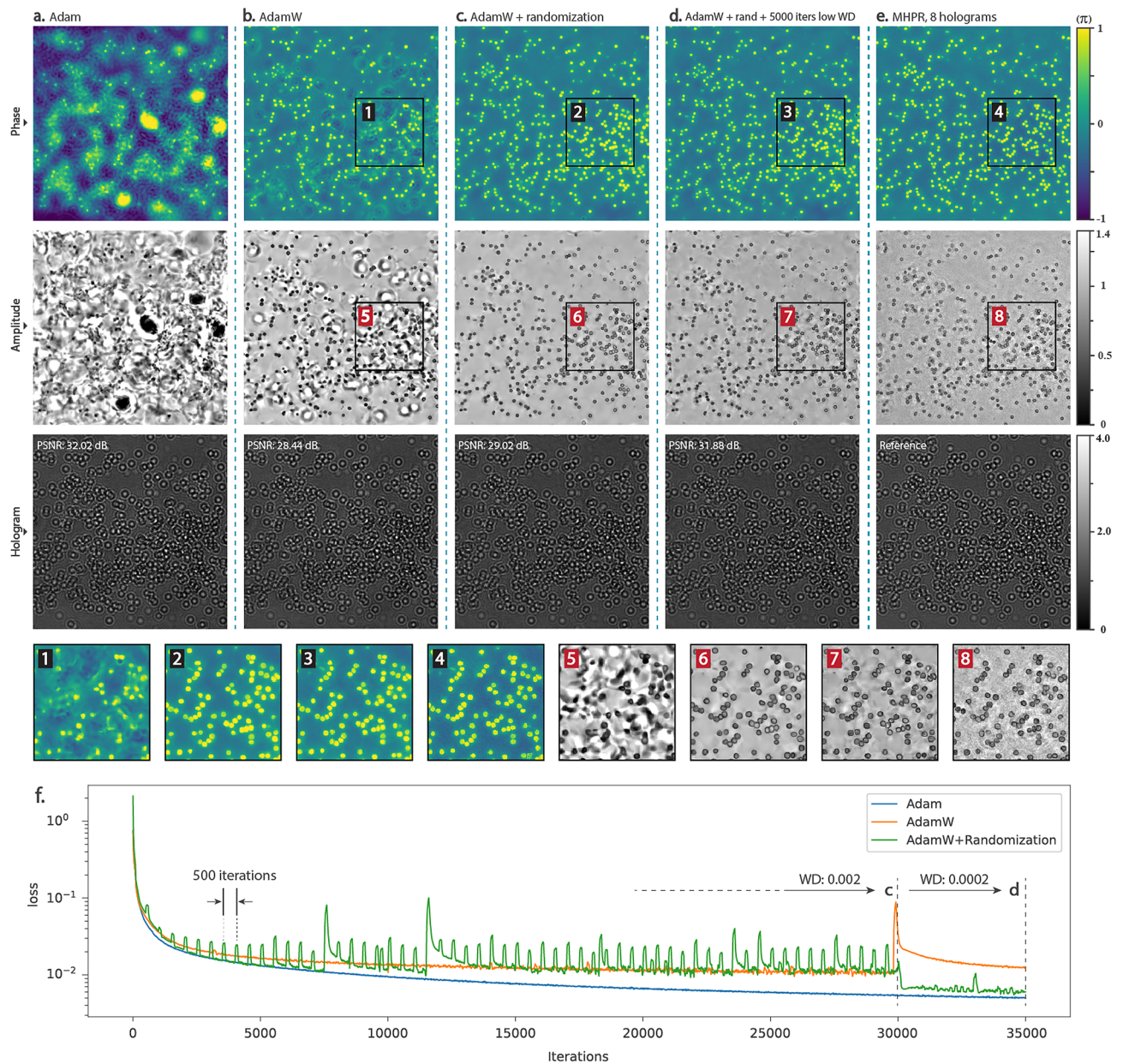


Figure 4. Reconstructed phase, amplitude, and diffraction (at the sensor plane) images of smeared red blood cells with different regularization settings. The results are obtained after 30,000 iterations (a) without any regularization, (b) with a weight decay of 0.002, (c) with randomizations applied every 500 iterations, and (d) with the same settings of (c) with additional 5000 iterations through which the weight decay is 0.0002 (no randomization). (e) The images obtained by the MHPR method using 8 holograms. Spikes on the green curve are caused by random perturbations.

showing any meaningful differences between the holograms and the object images. This effect is another consequence of overfitting.

Limitations

The problem of inverse reconstruction of an in-line hologram is underdetermined by (at least) a factor of 2; a real-valued intensity image against a complex-valued object function. Compressive sensing techniques address this problem by reducing the number of required parameters, assuming numerical constraints, and enforcing prior information about the object in the reconstruction process. For this reduction, it is required to transform the problem from the pixels space to the features space. Similarly, the DDN builds a pixel-wise estimate of the object field by a linear combination of feature maps. Hence, the performance of the reconstruction is limited by the level of indeterminism of the problem which is directly related to the amount of spatial (or spectral) details of the object, and the amount of optical information required to resolve through the reconstruction. These are two major limiting factors constraining the performance of the proposed method. The influences of these limitations will be explained by the following two experiments.

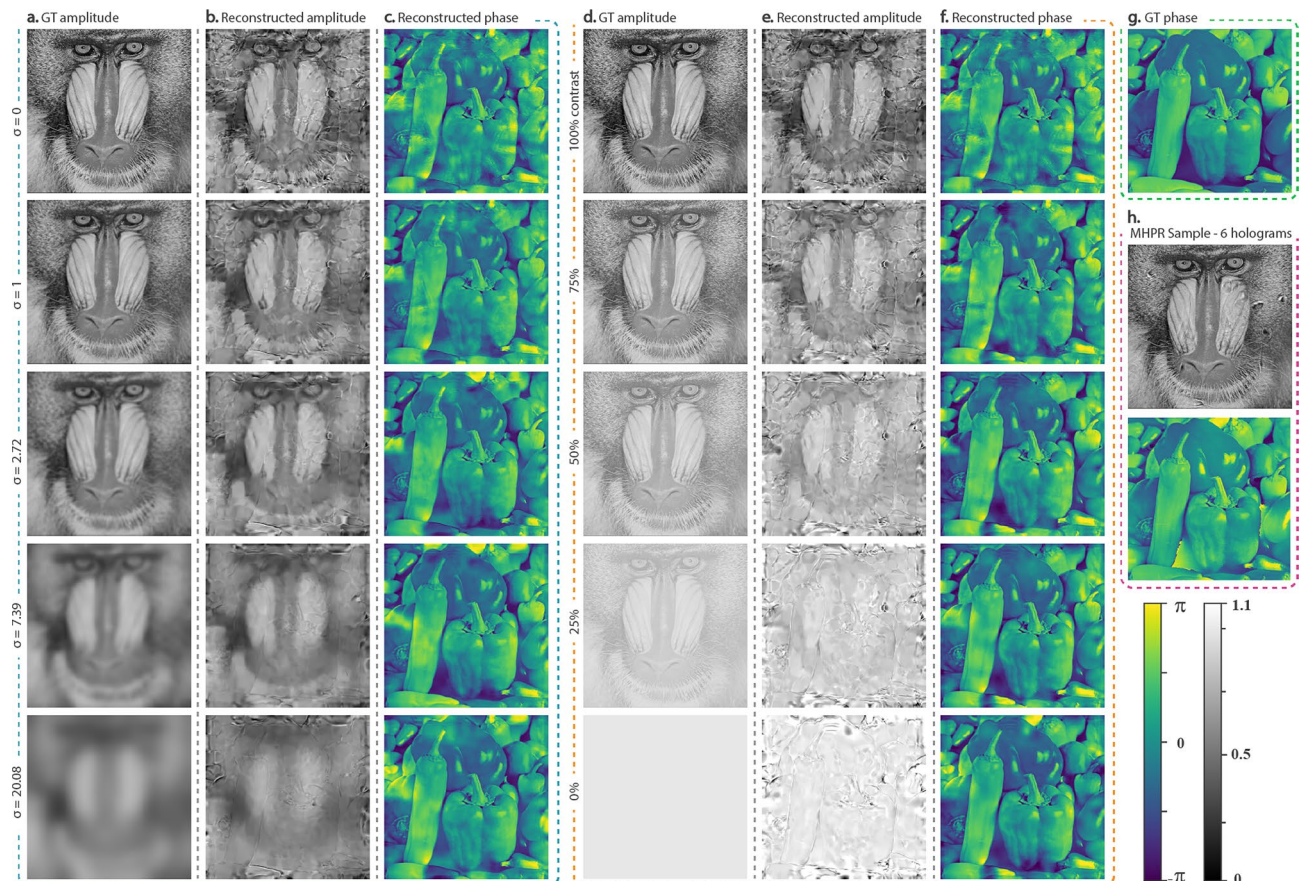


Figure 5. (a) Ground-truth (GT) amplitude images blurred by a Gaussian function with different degrees of standard deviation. (d) GT amplitude images with different contrasts. (b, c, e, f) Reconstructed images of (a, d, g) as the amplitude and phase images. (h) Reconstructed images achieved by MHPR method using 6 axially shifted holograms.

Level of ill-posedness. To investigate the effect of ill-posedness on the reconstruction performance, two general examples were considered in which, the effects of strength and distribution of details are examined. For both cases, five holograms were computationally generated by two randomly selected images as phase and amplitude.

In the first example, the amplitude image for each hologram is blurred by a Gaussian function with different degrees of smoothing power. The blurriness degree which is defined by a standard deviation parameter (σ) increases exponentially for each image such that $\sigma: \exp(0), \exp(1), \dots, \exp(3)$. The dimensions of the Gaussian kernels are defined by σ such that $d(\sigma) = 6 \times [\sigma] + 1$ where $[\cdot]$ is the round operation. This example gives us an insight into the influence of the distribution of spectral components on the ill-posedness of the problem and therefore, the accuracy of the reconstructions.

In the second example, five amplitude images with different contrasts are considered. The contrasts have a descending order from 100 to 0% with a 25% increment. This experiment is designed to investigate the relation between the power of spectral components of the object and the difficulty of the problem.

The input images and their reconstructions for both cases are shown in Fig. 5. The phase images in both experiments are unchanged for comparison.

The reconstructed images in Fig. 5 show that with any reduction of diversity and strength in details, the accuracy of the reconstructions improves. One explanation is that the Gaussian blur smooths the sharp edges of the amplitude image, which normally is the cause of ringing artifact; this effect is posed by the band limit in discrete Fourier analysis and is much more visible when phase and amplitude images are not geometrically correlated. Secondly, in some areas of the amplitude image, the pixel values are close to zero and because of finite sampling of digital image sensors, the phase information of those areas gets scrambled. Theoretically, these explanations are valid but as illustrated in Fig. 5h, these issues have minimum influence on the reconstructions when the available information about the object is enough.

To achieve a quantitative understanding of the results, we need to introduce several indices. The Normalized Root Mean Square (NRMS) index can be regarded as a measure of the scattering power of the holograms. NRMS of each hologram intensity image H can be calculated as:

Gaussian kernel STD*: σ	0	1	2.72	7.39	20.08
NRMS	0.71	0.71	0.72	0.72	0.71
PSE ($\times 10^3$)	4.0914	3.1272	2.0614	1.6437	1.4758
SSIM					
Amplitude	0.30	0.43	0.58	0.69	0.60
Phase	0.57	0.65	0.69	0.72	0.69

Table 1. Comparison of the SSIM and PSE indices of the reconstructed images in Fig. 5b,c respect to the scattering power of each simulated hologram. *STD standard deviation.

Contrast	100%	75%	50%	25%	0%
NRMS	0.71	0.63	0.55	0.52	0.59
PSE ($\times 10^3$)	4.0914	4.0915	4.0912	4.0923	–
SSIM					
Amplitude	0.30	0.26	0.19	0.11	0.65
Phase	0.57	0.65	0.71	0.72	0.71

Table 2. Comparison of the SSIM and PSE indices of the reconstructed images in Fig. 5e,f respect to the scattering power of each simulated hologram.

$$NRMS(\mathbf{H}) = \frac{RMS(\mathbf{H} - \mathbf{ref})}{RMS(\mathbf{ref})} \quad (6)$$

where \mathbf{ref} is the reference (or background) matrix. Moreover, the power spectral entropy (PSE)⁴⁹ of an image shows the information's uncertainty in a given spectrum which can be regarded as a measure of complexity. PSE of each amplitude image, \mathbf{I} , can be calculated as:

$$PSE(\mathbf{I}) = - \sum_{i=1}^n p_i \log_2 p_i \quad (7)$$

where p_i is the normalized PSE for each frequency component ω_i and can be described as:

$$p_i = p(\omega_i) = \frac{P(\omega_i)}{\sum_i P(\omega_i)} \quad (8)$$

P is the PSE of the image \mathbf{I} . Furthermore, the visual similarity between the reconstructed images and their ground-truth equivalents is commonly evaluated using the Structural Similarity Index Measure (SSIM). This quantitative measure of quality could be regarded as a measure of ill-posedness.

For both examples, these indices are calculated and arranged in Tables 1 and 2. For the first case, the NRMS values of the hologram images are showing that the information diversity in each amplitude image does not necessarily affect the scattering power of the hologram. The PSE values of the input amplitude images are showing a reduction of complexity with any increase of blurriness.

For the images in Fig. 5d, the NRMS of the simulated holograms have a descending behavior while the PSE of the amplitude images (except for zero contrast) is approximately constant. However, this opposite behavior of indices is followed by a descending behavior of the ill-posedness.

According to the SSIM values denoted in Table 1, information reduction directly results in improved reconstruction quality both in amplitude and phase. But for the case of contrast variation, another effect takes place as well. The SSIM values for the amplitude images denoted in Table 2 are implying the more obscured the details are, the more intensely the model filters them out, although the frequency components are the same. In other words, low-intensity details are regarded as noise and are more likely to be eliminated as can be seen in Fig. 5d. This behavior could be a result of shrinkage regularization of weight decay which inclines the model to an underfitting situation. The opposite slopes of variations in the SSIM values in Table 2 is a quantitative appearance of this behavior.

Regarding the scattering power of the hologram or complexity of the amplitude in each image set, the level of ill-posedness of the problem controls the performance of the reconstruction. This level increases when the specimen is thick or has any axial distribution. Hence, it is not surprising to say that the model performs better when resolving sparsely distributed details in a smooth and well-illuminated background which is the case of thin microscopic biological samples.

Sparsity-fidelity tradeoff. When the object function is seen as a set of features, the resolving power of the model would depend on the geometrical characteristics of the details. When regularization and under-param-

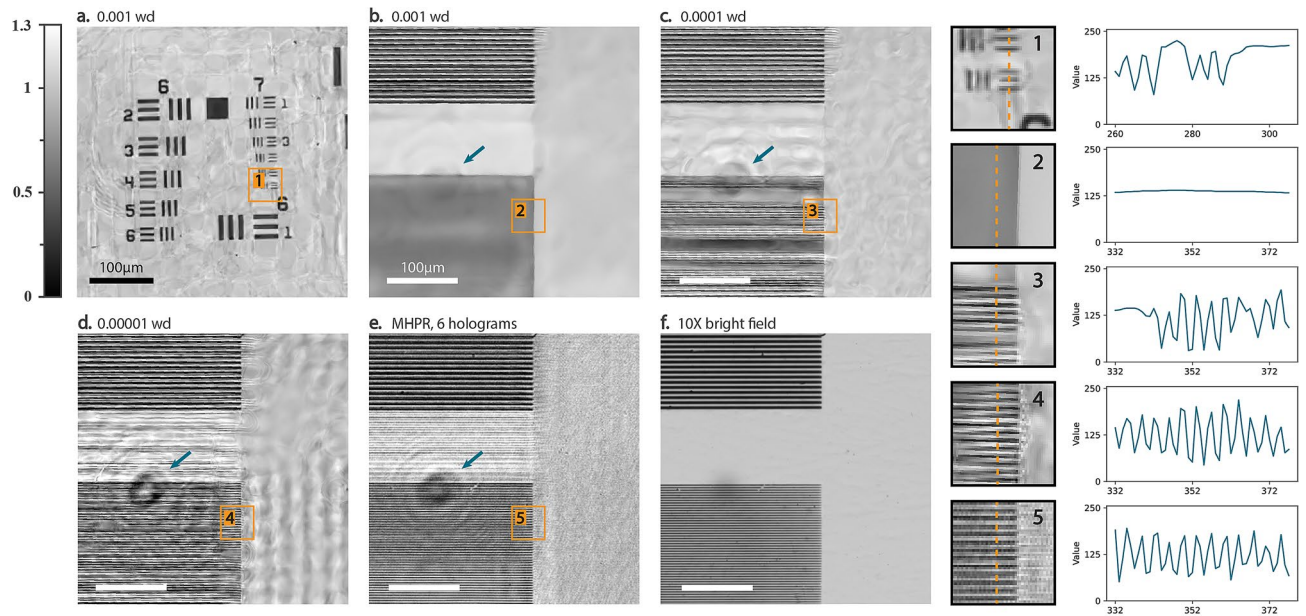


Figure 6. (a) the reconstructed amplitude of a standard resolution target (228 lp/mm for the finest details). (b, c, d) The amplitude images recovered from the transmissive grating (125 and 250 lp/mm) by DCOD with weight decay ranging from 10^{-3} to 10^{-5} . (e) The reconstructed amplitude image obtained by multi-height phase recovery method using 6 holograms. (f) The microscopic image captured from the same field of view using a 10×0.2 N.A. objective lens.

eterization are applied simultaneously, it is unlikely to achieve the maximum resolution for every geometrical feature. To visualize this relationship between the geometrical characteristics of the object and the resolving power of the model, two types of samples are provided: a USAF 1951 standard resolution target, and a transmissive grating. The reconstructed amplitude images of these samples are shown in Fig. 6. The grating has 125 and 250 line pairs per millimeter while the finest section in the resolution target has 228 lp/mm which approximately has the same linewidth of $\sim 2\ \mu\text{m}$.

The finest details of the resolution sample in Fig. 6a are accurately resolved while -with the same model settings- the fine lines in the grating sample are averaged as a smooth area. Fine details in both samples require the same resolving power but the model prioritizes small, isolated features and ignores pattern-like features as noise. As illustrated in Fig. 6c,d, the fine grating pattern can be recovered by decreasing the weight decay, but it comes with the expense of amplified noise.

Different phenomena affect the resolution of the reconstructions e.g., SNR (signal-to-noise ratio), or information diversity i.e., the amount of information acquired under different experimental conditions. The information diversity for our experiments is constant but the under-sampling causes ill-posedness as examined in the previous section. Furthermore, with sufficient exposure to illumination, the contribution of environmental noises tends to zero while the coherent-related noises persistently keep their contribution in the signal. While these issues are consistent in the problem, the fine details could be resolved with lower regularization ratios as shown in Fig. 6c,d. Hence, the smoothing effect in Fig. 6b cannot be related to the noises or information diversity. This effect is a consequence of a fundamental tradeoff between sparsity and fidelity; the fine lines in the resolution sample are distributed much sparser than in the grating.

Weight decay regularization does not impose sparsity but reduces the variations of the features obtained from the object. The variation reduction occurs in the parameter space under the influence of the under-parameterization of the network which results in the extraction of a simplified representation of the object in the feature domain. Weight decay further penalizes the parameters to gather as few varying features as possible, thus enforces a sparsity constraint on the feature extraction process. Sparsity promotion in the proposed method manifests as a smoothing effect over large spatial features and its strength can be controlled by the weight decay coefficient. In general terms, this effect improves the overall quality of the reconstructions by removing high-frequency noises especially the artifacts originating from other sources than the object plane e.g., the fringes formed by an out-of-focus object marked on the images in Fig. 6.

Outlook. In this study, we demonstrated that an untrained regularized convolution neural network can iteratively recover the object information that is partially available in an in-line hologram. Unlike other methods, this operation does not need to be trained on any dataset or to be supervised by any excess information about the object. The regularization induced by DIP, under-parameterization, weight decay, and randomization that are employed through our demonstration is general and does not need any specific tuning per specimen or external supervision. However, all these advantages come with the price of requiring extensive computational resources and long processing times. Additionally, the trade-off between the resolving power and the noise seems to be

fundamental and the problem of limited phase range due to lack of any phase recovery constraint has remained unsolved.

Since the twin image noise typically scrambles phase information, the commonly used deterministic phase unwrapping techniques are usually not helpful in in-line holography. Also, the proposed model fails to reconstruct the wrapped phase correctly and presents random artifacts instead. To overcome this issue, one straight forward way is performing more measurements. With just one more measurement under a different condition, we can obtain an unwrapped estimate of the phase. For instance, if the transport-of-intensity equation (TIE) is applied to several images taken from different heights, it provides an initial phase guess⁶. After that, a regularization term can be defined which will favor achieving a minimized smooth difference between the output phase and the guess. Another alternative idea is using a supervised model to give an unwrapped initial guess of the phase profile. After feeding the results to a fidelity term, the model can be regularized to reconstruct unwrapped phase information. Another possible solution is obtaining multiple phase images through multiple channels with uniquely designed constraints each of which constructing a portion of phase information.

Speed issues can be solved by tuning hyperparameters, adopting an optimized network architecture, or making use of advanced algorithms for randomization, such as warm restart, for greater efficiency. Initializing network parameters (especially the input random tensor) with an encoder unit in a supervised or unsupervised manner might be beneficial as well.

We believe the DCOD structure is versatile enough to be modified to solve other inverse problems with little information about the signal or limited experimental resources such as tomographic imaging using holographic or cross-sectional images, or quantitative phase imaging using a set of intensity images.

Materials and methods

Experimental scheme. A lensless in-line holographic microscope is designed and arranged as schematically shown in Fig. 7. The image sensor is a CMOS Sony IMX-219 which is driven by a Raspberry Pi version 2 camera module. This RGB image sensor provides an array of 10-bit depth (16-bit digital array) 3296×2480 pixels per channel with $1.12 \mu\text{m}$ pixel pitch which gives an effective area of $3692 \mu\text{m} \times 2778 \mu\text{m}$ as the field of view for in-line holographic imaging. A Raspberry Pi 3 B mini-computer board controls the camera system to capture, preprocess, store, and stream the images.

The illumination source includes a 40 mW (max) 532.3 nm Nd:YAG laser with ~ 1 nm bandwidth. The operational power of the laser is reduced to less than 10% for practical use. The beam is spatially filtered by passing through a $20 \mu\text{m}$ pinhole. The pinhole is also placed ~ 10 cm away from the sample plane which is far enough to ensure the illumination wavefront is planar compared to the microscopic samples of size $< 1000 \mu\text{m}$. A 3.2 cm focal length lens is placed before the spatial filter to provide a broader thus more uniform illumination on the sensor plane. The sample slide is placed near the sensor plane ($200 \mu\text{m} - 2$ mm) to minimize the effect of undesired cross-interference. Moreover, acquiring additional images at different heights for MHPR could pose undesired lateral displacement against any small variations of height. For small sample-to-sensor distances, this effect could be minimized. With these settings, the illumination light has temporal coherence for a length of $\sim 283 \mu\text{m}$ (in a vacuum) and spatial coherence for a radius of $\sim 847 \mu\text{m}$ on the sample plane which is more than enough for typical microscopic samples. This feature of small *sample-to-sensor* (z_s)/*sample-to-source* (z_i) ratio allows achieving unit fringe magnification with a resolution as small as the pixel size of the sensor (if $> \lambda/2$) and with a field of view as large as the area of the sensor, resulting in a large spatial-bandwidth product.

Image acquisition. For each experiment, 6 holograms are captured from a sequence of heights respectively (with $\sim 100 \mu\text{m}$ interval) due to validation via the MHPR algorithm (Supplementary section B). The hologram with the smallest sample-to-sensor distance is considered as the reference hologram to be adopted by our approach.

Preprocessing. The initially recorded raw 10bit Bayer image is needed to demosaic at the beginning. Then, the green channel of the outcoming RGB image is assigned as the hologram. A 512×512 pixels region of each hologram is cropped to feed the model for processing.

To bring quantitative consistency to the measurements, each recorded image is divided with a pre-acquired background image captured in a similar experimental setting.

Network design. For network M , a DDN is designed to map a stack of k_0 many n_0 -dimensional random matrices as a tensor $\mathbf{B}_0 \in \mathbb{R}^{n_0 \times k_0}$, to a n_d -dimensional double-channeled ($k_{out} = 2$) image. In each layer, the transformation process includes an element-wise linear combination of the channels multiplied by a weight tensor followed by upsampling, normalization, and regularization by a rectified linear unit (ReLU). We can write the operations in each given $(i + 1)$ -th layer as:

$$\mathbf{B}_{i+1} = cn(\text{relu}(u_i \mathbf{B}_i \mathbf{w}_i)), \quad i = 0, \dots, d - 1 \quad (9)$$

$\mathbf{w}_i \in \mathbb{R}^{k_i \times k_{i+1}}$ is the weight tensor of each layer and u_i is a bi-linear upsampling operator. $cn(\cdot)$ performs a channel normalization operation (known as batch normalization) that specifically for its input tensor $\mathbf{Z}_i = \text{relu}(u_i \mathbf{B}_i \mathbf{w}_i)$, for each channel j , gives:

$$\mathbf{Z}'_{ij} = \frac{\mathbf{Z}_{ij} - \text{mean}(\mathbf{Z}_{ij})}{\text{std}(\mathbf{Z}_{ij})} \gamma_{ij} + \beta_{ij} \quad (10)$$

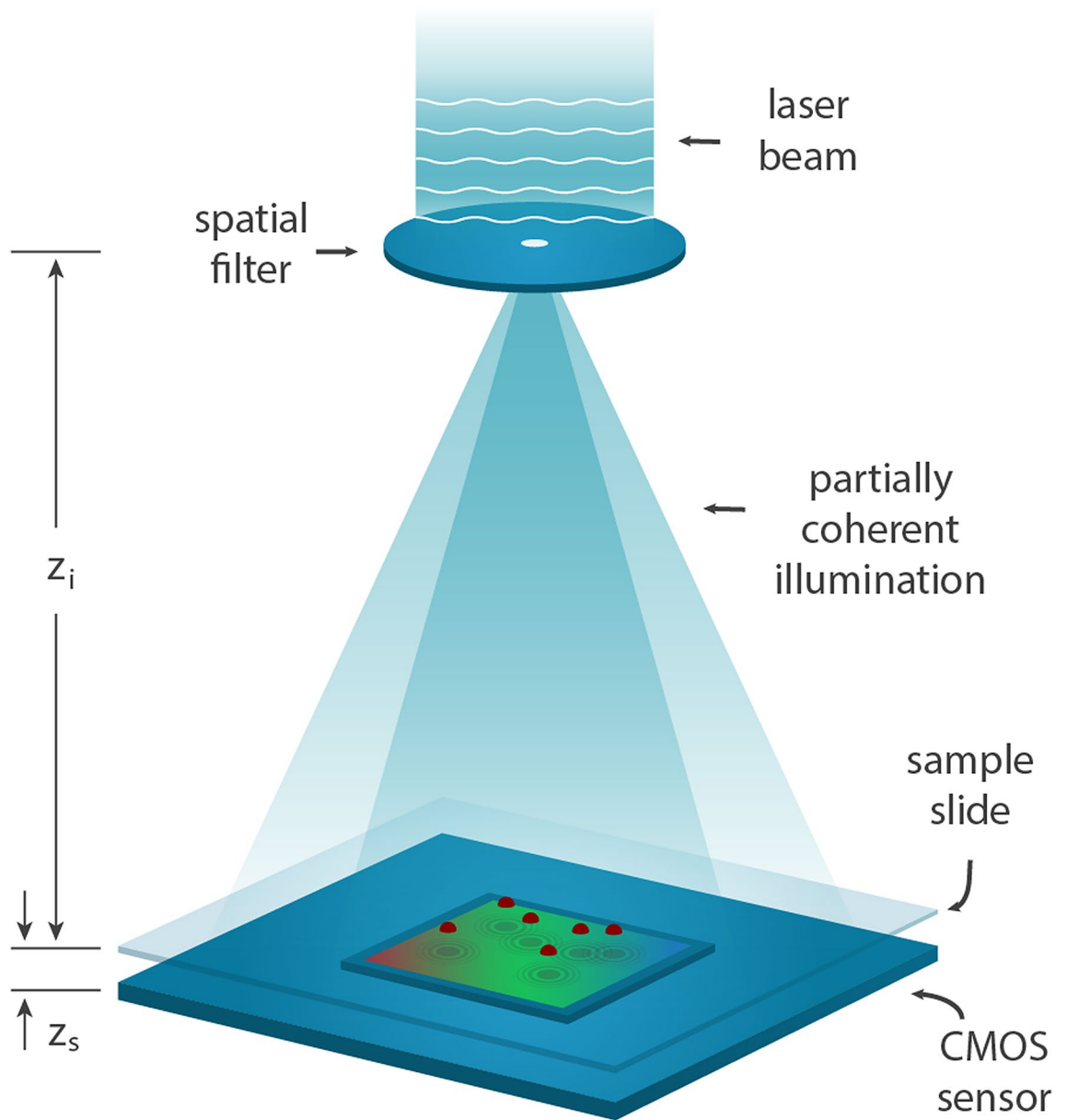


Figure 7. Sketch of the experimental setup for $z_i \simeq 10$ cm and $z_s < 2$ mm.

where *mean* and *std* are empirical mean and standard deviation and γ_{ij} and β_{ij} are learnable parameters³⁸. Subsequently, the output of the d -layer network would be:

$$\{\phi^*, \mathbf{A}^*\} = 2\pi \text{sigmoid}(\mathbf{B}_d \mathbf{w}_d), \quad \mathbf{w}_d \in \mathbb{R}^{k_d \times 2} \quad (11)$$

which models amplitude \mathbf{A}^* and phase ϕ^* images through two output channels of M . Based on the layout proposed in³⁸, the default network is constructed by 5 similar layers with 256 channels, twofold bi-linear upsampling layers, and an output layer with the same setting but without upsampling. In each layer, a 1×1 convolution operation performs both weight tensor multiplication and linear combination operations as expressed in Eq. (11). The network is finally fed by a random tensor (\mathbf{B}_0) with $n_0 = 16 \times 16$ and $k_0 = 256$ which results in a two channeled $n_d = 512 \times 512$ phase-intensity pair tensor. For more details about this network structure, see Supplementary information section D. This framework is implemented using the TensorFlow version 2.3.0 platform in Python 3.6.9.

Training. The fixed $16 \times 16 \times 256$ input tensor contains normally distributed random values with 0 mean and 0.1 standard deviation. In all experiments (except those mentioned particularly in their descriptions), the models are trained with a learning rate of 0.01 and weight decay of 0.002. For weight decay regularization, the TensorFlow implementation of AdamW is adopted from.

With fixed hyperparameters, (for most samples) the proposed model needs about 30,000 iterations to produce satisfactory results which on an Nvidia Tesla k80 GPU, the process takes ~ 40 min.

Randomization. Every 500 iterations:

1. A Gaussian noise tensor with a mean of 0 and a standard deviation of 0.02 should be added to the input tensor.
2. The amplitude image should be multiplied by a coefficient whose value switches between 1.3 and 1.4 in every randomization step.

Regardless of randomization, the amplitude image must be multiplied by a coefficient larger than 1, although the amplitude of every pixel is expected to be less than 1 unless the model cannot generate satisfactory results. In general, choosing every value for this coefficient between 1.1 and 1.8 does not show any clear impact on the reconstruction performance.

Simulations. For the simulations shown in Fig. 5, the object plane is considered at $z = 300 \mu\text{m}$, the illumination is assumed to be coherent, planar, and has a wavelength of 532.2 nm. Additionally, the pixels of the holograms are assumed to have a size of $1.12 \mu\text{m}$. For the MHPR method, the height differences are similar and equal to $50 \mu\text{m}$. The planes are considered further away from the reference plane which is at $z = 300 \mu\text{m}$.

Samples.

1. Unlabeled de-identified and existing dry Oral epithelial smear slides (cheek cells) and Blood smear slides are obtained from the Microfluidics Laboratory at SBU Laser and Plasma Research Institute.
2. The 1951 USAF standard resolution target is purchased from Thorlabs (# R3L3S1P).
3. The Transmissive grating is fabricated by photolithography technique applied on a copper-coated glass substrate and is obtained from the Surface and Material Laboratory at SBU.

Data availability

The images and codes are publicly available in <https://github.com/farhadnkm/DCOD>.

Received: 5 February 2021; Accepted: 10 May 2021

Published online: 25 May 2021

References

1. Mudanyali, O. *et al.* Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications. *Lab. Chip* **10**, 1417–1428 (2010).
2. Molony, K. M., Hennelly, B. M., Kelly, D. P. & Naughton, T. J. Reconstruction algorithms applied to in-line Gabor digital holographic microscopy. *Opt. Commun.* **283**, 903–909 (2010).
3. Yu, X., Hong, J., Liu, C. & Kim, M. K. Review of digital holographic microscopy for three-dimensional profiling and tracking. *Opt. Eng.* **53**, 1–22 (2014).
4. Wu, Y. & Ozcan, A. Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring. *Methods* **136**, 4–16 (2018).
5. Rivenson, Y. *et al.* Sparsity-based multi-height phase recovery in holographic microscopy. *Sci. Rep.* **6**, 37862 (2016).
6. Greenbaum, A. *et al.* Wide-field computational imaging of pathology slides using lens-free on-chip microscopy. *Sci. Transl. Med.* **6**, 267ra175 (2014).
7. Goodman, J. W. Introduction to Fourier optics. In *Introduction to Fourier Optics* 3rd edn, Vol. 1 (ed. Goodman, J. W.) (CO Roberts Co. Publ., 2005).
8. Cucho, E., Marquet, P. & Depeursinge, C. Spatial filtering for zero-order and twin-image elimination in digital off-axis holography. *Appl. Opt.* **39**, 4070–4075 (2000).
9. Yamaguchi, I. & Zhang, T. Phase-shifting digital holography. *Opt. Lett.* **22**, 1268–1270 (1997).
10. Cucho, E., Bevilacqua, F. & Depeursinge, C. Digital holography for quantitative phase-contrast imaging. *Opt. Lett.* **24**, 291–293 (1999).
11. Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758–2769 (1982).
12. Gerchberg, R. W. & Saxton, W. O. Practical algorithm for determination of phase from image and diffraction plane pictures. *Optik (Stuttgart)* **35**, 237 (1972).
13. Fienup, J. R. & Wackerman, C. C. Phase-retrieval stagnation problems and solutions. *J. Opt. Soc. Am. A* **3**, 1897–1907 (1986).
14. Greenbaum, A. & Ozcan, A. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Opt. Express* **20**, 3129–3143 (2012).
15. Zhang, Y., Pedrini, G., Osten, W. & Tiziani, H. J. Whole optical wave field reconstruction from double or multi in-line holograms by phase retrieval algorithm. *Opt. Express* **11**, 3234–3241 (2003).
16. Bao, P., Situ, G., Pedrini, G. & Osten, W. Lensless phase microscopy using phase retrieval with multiple illumination wavelengths. *Appl. Opt.* **51**, 5486–5494 (2012).
17. Luo, W., Greenbaum, A., Zhang, Y. & Ozcan, A. Synthetic aperture-based on-chip microscopy. *Light Sci. Appl.* **4**, e261–e261 (2015).
18. Camacho, L., Micó, V., Zalevsky, Z. & García, J. Quantitative phase microscopy using defocusing by means of a spatial light modulator. *Opt. Express* **18**, 6755–6766 (2010).
19. Micó, V., García, J., Zalevsky, Z. & Javidi, B. Phase-shifting Gabor holography. *Opt. Lett.* **34**, 1492–1494 (2009).
20. Elser, V. Phase retrieval by iterated projections. *J. Opt. Soc. Am. A* **20**, 40–55 (2003).
21. Marchesini, S. Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Rev. Sci. Instrum.* **78**, 11301 (2007).
22. Latychevskaia, T. & Fink, H.-W. Solution to the twin image problem in holography. *Phys. Rev. Lett.* **98**, 233901 (2007).

23. Denis, L., Lorenz, D., Thiébaud, E., Fournier, C. & Trede, D. Inline hologram reconstruction with sparsity constraints. *Opt. Lett.* **34**, 3475–3477 (2009).
24. Lim, S., Marks, D. L. & Brady, D. J. Sampling and processing for compressive holography Invited. *Appl. Opt.* **50**, H75–H86 (2011).
25. Brady, D. J., Choi, K., Marks, D. L., Horisaki, R. & Lim, S. Compressive holography. *Opt. Express* **17**, 13040–13049 (2009).
26. Jolivet, F. *et al.* Regularized reconstruction of absorbing and phase objects from a single in-line hologram, application to fluid mechanics and micro-biology. *Opt. Express* **26**, 8923–8940 (2018).
27. Goy, A., Arthur, K., Li, S. & Barbastathis, G. Low photon count phase retrieval using deep learning. *Phys. Rev. Lett.* **121**, 243902 (2018).
28. Lyu, M. *et al.* Deep-learning-based ghost imaging. *Sci. Rep.* **7**, 17865 (2017).
29. Lyu, M., Wang, H., Li, G., Zheng, S. & Situ, G. Learning-based lensless imaging through optically thick scattering media. *Adv. Photonics* **1**, 1–10 (2019).
30. Ren, Z., Xu, Z. & Lam, E. Y. Learning-based nonparametric autofocusing for digital holography. *Optica* **5**, 337–344 (2018).
31. Wang, K., Li, Y., Kemao, Q., Di, J. & Zhao, J. One-step robust deep learning phase unwrapping. *Opt. Express* **27**, 15100–15115 (2019).
32. Rivenson, Y., Zhang, Y., Günaydin, H., Teng, D. & Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).
33. Wang, H., Lyu, M. & Situ, G. eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Opt. Express* **26**, 22603–22614 (2018).
34. Jo, Y. *et al.* Quantitative phase imaging and artificial intelligence: a review. *IEEE J. Sel. Top. Quantum Electron.* **25**, 1–14 (2019).
35. Wang, F. *et al.* Phase imaging with an untrained neural network. *Light Sci. Appl.* **9**, 77 (2020).
36. Bostan, E., Heckel, R., Chen, M., Kellman, M. & Waller, L. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. *Optica* **7**, 559–562 (2020).
37. Ulyanov, D., Vedaldi, A. & Lempitsky, V. Deep image prior. *Int. J. Comput. Vis.* **128**, 1867–1888 (2020).
38. Heckel, R. & Hand, P. Deep decoder: concise image representations from untrained non-convolutional networks. Preprint at <https://arxiv.org/abs/1810.03982> (2019).
39. Latychevskaia, T. & Fink, H.-W. Practical algorithms for simulation and reconstruction of digital in-line holograms. *Appl. Opt.* **54**, 2424–2434 (2015).
40. Kim, M. K. Principles and techniques of digital holographic microscopy. *SPIE Rev.* **1**, 1–51 (2010).
41. Ersoy, O. K. *Diffraction, Fourier Optics and Imaging* (Wiley, New York, 2007).
42. Bioucas-Dias, J. M. & Figueiredo, M. A. T. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **16**, 2992–3004 (2007).
43. Wu, J. *et al.* Single-shot lensless imaging with fresnel zone aperture and incoherent illumination. *Light Sci. Appl.* **9**, 53 (2020).
44. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2017).
45. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. Preprint at <https://arxiv.org/abs/1711.05101> (2019).
46. Krogh, A. & Hertz, J. A. A simple weight decay can improve generalization. In *Advances in Neural Information Processing Systems* Vol. 4 (eds Moody, J. E. *et al.*) 950–957 (Morgan-Kaufmann, London, 1992).
47. Duchi, J., Hazan, E. & Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **12**, 2121–2159 (2011).
48. Reddi, S. J., Kale, S. & Kumar, S. On the convergence of Adam and beyond. Preprint at <https://arxiv.org/abs/1904.09237> (2019).
49. Pan, Y. N., Chen, J. & Li, X. L. Spectral entropy: a complementary index for rolling element bearing performance degradation assessment. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **223**, 1223–1231 (2009).

Author contributions

F.N. conceptualized the idea, designed the approach, and carried out programming and computations. F.N., H.Q., and H.L. conducted experiments and data preparation and wrote the manuscript. H.L. supervised the research and provided funding and research supplies. All the authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-90312-5>.

Correspondence and requests for materials should be addressed to F.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021