جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Holographic 3D particle imaging with model-based deep network

Item Type	Article
Authors	Chen, Ni;Wang, Congli;Heidrich, Wolfgang
Citation	Chen, N., Wang, C., & Heidrich, W. (2021). Holographic 3D particle imaging with model-based deep network. IEEE Transactions on Computational Imaging, 1–1. doi:10.1109/tci.2021.3063870
Eprint version	Publisher's Version/PDF
DOI	10.1109/TCI.2021.3063870
Publisher	Institute of Electrical and Electronics Engineers (IEEE)
Journal	IEEE Transactions on Computational Imaging
Rights	This work is licensed under a Creative Commons Attribution 4.0 License.
Download date	2024-04-18 20:53:53
Item License	https://creativecommons.org/licenses/by/4.0/
Link to Item	http://hdl.handle.net/10754/666339

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

Holographic 3D particle imaging with model-based deep network

Ni Chen, Congli Wang, and Wolfgang Heidrich, *Fellow, IEEE*

***Abstract*— Gabor holography is an amazingly simple and effective approach for three-dimensional (3D) imaging. However, it suffers from a DC term, twin-image entanglement, and defocus noise. The conventional approach for solving this problem is either using an off-axis setup, or compressive holography. The former sacrifices simplicity, and the latter is computationally demanding and time-consuming. To cope with this problem, we propose a model-based holographic network (MB-HoloNet) for three-dimensional particle imaging. The free-space point spread function (PSF), which is essential for hologram reconstruction, is used as a prior in the MB-HoloNet. All parameters are learned in an end-to-end fashion. The physical prior makes the network efficient and stable for both localization and 3D particle size reconstructions.**

***Index Terms*— Holography, Neural networks, Inverse problems**

I. Introduction

Holography is a powerful tool for 3D imaging and display due to its wavefront encoding ability. Since it has been proposed in 1948 [1], a large amount of work has emerged for solving the DC and twin-image problems in the hologram reconstruction. These problems can be improved by changing the optical system, such as off-axis holography [2], phase-shifting holography [3], optical scanning holography [4], and others. However, the simplicity and high space-bandwidth product of Gabor holography remains attractive and competitive. Apart from the DC term and twin-image, the main issue in 3D digital holographic imaging is the defocus noise. Compressive sensing approaches [5]–[7] are efficient in restoring 3D images from holograms. However, these methods are computationally intensive and usually require fine-tuned parameters, such as the regularization and relaxation ones [5], [8].

Recently, deep learning has emerged as a comprehensive tool in computational imaging [9], in areas of optical tomography [10], ghost imaging [11], digital holography [12], imaging through scattering media [13], and phase imaging [14]. However, compared to other fields, deep learning has been underutilized in digital holography [15], especially for 3D holographic imaging. Most previous works for optical imaging rely on an end-to-end neural network, which requires large amounts of data for training. 3D holographic imaging is challenging for data-driven networks due to the limitations of acquiring large quantities of training data. There are mainly two popular ways for obtaining the training data-sets: (1) Display the target (labels) on spatial light modulators (SLM) or digital micromirror devices (DMD) for capturing the images (data) of the labels that passing through the optical system [11]; (2) Capture a few high-resolution optical images, calculate the labels with conventional imaging methods. More data and labels are obtained by cropping, rotating of the existing ones [16]. Both approaches are not practical for 3D images. Therefore, for 3D holographic imaging, only a few results have been reported [17]–[19]. The state-of-the-art technique can reconstruct particle locations in a volume [17], [18]. The authors obtained training data with a previously well-developed technique, for example, scanning microscopy, and a modified U-Net was applied for particle localization requiring time-consuming measurements (e.g. scanning microscopy) to obtain ground-truth particle training datasets.

Besides holography, many optical imaging techniques rely on specific physical models. Most of the previous works rely on an end-to-end neural network while neglecting the knowledge in established physics models [20]. More specifically, while the initial estimates are physically accurate, the network architectures are not inherently physics-aware. As such there are no guarantees that the final result will indeed be physically accurate. To integrate physical knowledge as network priors, several model-based schemes have been proposed to incorporate priors with the forward model in the learning process to ensure data consistency and prior suitability, known as the unrolled networks [21]–[23]. Unrolled networks mimic the behavior of traditional optimization-based solvers, since each stage corresponds to one iteration of such an optimization. Furthermore, physics constraints can be explicitly enforced in each stage/iteration, making the whole network inherently physics-aware. Such combined approaches have been successfully applied to computed tomography (CT) [24] and magnetic resonance imaging (MRI) reconstruction [25]. In particular, Kristina et al. [25] have unrolled a traditional model-based optimization algorithm based on the alternating direction method of multipliers (ADMM) [26] with a convolution neural network (CNN) for mask-based lens-less imaging.

Inspired by previous studies, we integrate the physical principles of holography with deep learning-based approaches for 3D holographic particle imaging. We consider 3D holographic imaging as a compressive sensing problem and adopting an iterative shrinkage-thresholding algorithm to perform 3D hologram reconstruction, named model-based holonet (MB-

Compiled on February 24, 2021. This work was supported by baseline funding of the King Abdullah University of Science and Technology.

N. Chen, C. Wang, and W. Heidrich are with the Visual Computing Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia (e-mails: {ni.chen, congli.wang, wolfgang.heidrich}@kaust.edu.sa).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

HoloNet). In particular, MB-HoloNet takes a single Gabor hologram along with the free-space PSF and outputs the 3D particle volume. It is composed of a fixed number of stages, each of which strictly corresponds to an iteration in the shrinkage-thresholding algorithm. Features of the objects and all the parameters involved are learned in an end-to-end fashion. That said, MB-HoloNet differentiates the algorithmic parameters with the labor heavy parameter tuning done in the network training process. Compared to prior arts, our proposed MB-HoloNet leverages both the physical models and the data-driven learning techniques, in a differentiable unified optimization framework. The PSF fed into the network makes it more accurate and numerically stable. MB-HoloNet can work when the test holograms are captured under different environments, as in the training data. We further show that the network trained with synthesized data works for experimental captured hologram reconstruction, as a natural consequence of the physically motivated property of our proposed learning framework. We believe MB-HoloNet will serve as an initial point for future unrolled networks for holographic reconstruction.

II. Physically based MB-HoloNet

The proposed method consists of two steps: (i) Linearization of Gabor hologram imaging; (ii) Solving the linear problem augmented with data-driven priors in MB-HoloNet.

A. Linearization of Gabor holographic imaging

In a Gabor holographic imaging system, as in Fig. 1, a 3D object of amplitude transfer function $o(\mathbf{r}, z)$ is positioned at the origin of the coordinate system, with $\mathbf{r} = (x, y)$. A plane reference wave field $\exp(jkz)$ of wave number $k = 2\pi/\lambda$ illuminates the object, producing the object field $o(\mathbf{r}, z)\exp(jkz)$, with position z encoded in the phase delay.

Both the reference and the object field propagates along the z-axis, arriving at the sensor plane at z_0. According to scalar diffraction theory [27], with $h(\mathbf{r}, z)$ denoting free-space propagation of point $(\mathbf{r}, z)$ and $\otimes$ for convolution, the impinging 2D fields are $u_r(\mathbf{r}) \approx 1$ and

$$u_o(\mathbf{r}) = \int_{z_1}^{z_2} \Big(o(\mathbf{r}, z)\exp(jkz)\Big) \otimes h(\mathbf{r}_h - \mathbf{r}, z_0 - z)\,\mathrm{d}z. \quad (1)$$

The two fields u_r and u_o interfere with each other and produce an interferometric intensity pattern $I_h = |u_r + u_o|^2$, known as the Gabor hologram, and is captured by the image sensor. Expanding I_h with $\Re\{\cdot\}$ denoting the real part and $\overline{\{\cdot\}}$ denoting the complex conjugate:

$$I_h(\mathbf{r}) = 2\Re\{u_o\overline{u}_r\} + |u_r|^2 + |u_o|^2, \quad (2)$$

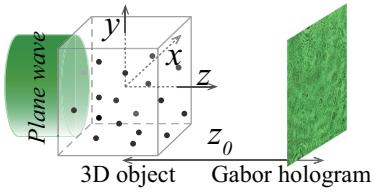


Fig. 1: Schematic image of an in-line holography system.

where $|u_r|^2$ is known from calibration, and $|u_o|^2$ is treated as a non-linear model error. With these, the hologram reconstruction is rephrased as a linear inverse problem:

$$\mathbf{I}_h \approx 2\Re\{\mathbf{P}\mathbf{o}\}, \quad (3)$$

where bold fonts denote vectors in linear algebra, discretized from their continuous versions, and $\mathbf{P}$ is a linear operator that maps 3D complex fields to 2D complex fields.

B. Model-based HoloNet

Due to the large solution space of $\mathbf{o}$, it is ill-posed for directly solving Eq. (4). By assuming that $\mathbf{o}$ is real valued with physically realistic sparsity properties [28], [29] in linear basis $\mathbf{\Psi}$, regularized least squares fitting Eq. (3) yields:

$$\min_{\mathbf{o}} \quad \frac{1}{2}\|\mathbf{\Phi}\mathbf{o} - \mathbf{I}_h\|_2^2 + \gamma\|\mathbf{\Psi}\mathbf{o}\|_1, \quad (4)$$

where $\mathbf{\Phi} = 2\Re\{\mathbf{P}\}$, $\|\cdot\|_2^2$ is the ℓ_2-norm squared, expressing fitting fidelity between measurement and the physical model, $\|\cdot\|_1$ is the ℓ_1-norm, and γ is a trade-off parameter. Though coupled, the two terms in Eq. (4) can be de-coupled and alternatively solved via the half-quadratic splitting method [30]. By introducing an additional variable $\mathbf{v} = \mathbf{o}$, the two terms are separated:

$$\min_{\mathbf{o},\mathbf{v}} \quad \frac{1}{2}\|\mathbf{\Phi}\mathbf{o} - \mathbf{I}_h\|_2^2 + \gamma\|\mathbf{\Psi}\mathbf{v}\|_1 + \mu\frac{1}{2}\|\mathbf{o} - \mathbf{v}\|_2^2, \quad (5)$$

with $\mu \to \infty$ being an algorithmic parameter. At the n^{th} iteration, solving $\mathbf{o}$ and $\mathbf{v}$ yields the following two update steps in sequence:

$$\mathbf{o}^{n+1} = \arg\min_{\mathbf{o}} \quad \|\mathbf{\Phi}\mathbf{o} - \mathbf{I}_h\|_2^2 + \mu\|\mathbf{o} - \mathbf{v}^n\|_2^2, \quad (6)$$

$$\mathbf{v}^{n+1} = \arg\min_{\mathbf{v}} \quad \frac{1}{2}\|\mathbf{v} - \mathbf{o}^n\|_2^2 + \frac{\gamma}{\mu}\|\mathbf{\Psi}\mathbf{v}\|_1. \quad (7)$$

Equation (6) and Eq. (7) can be solved efficiently using proximal algorithms [26], [31]. However, these methods usually require hundreds of iterations to obtain a satisfactory result, which inevitably gives rise to high computational cost and is thus restricting for fast applications. In addition, all parameters must be pre-defined, and it can be quite challenging to tune them priori. Also, the selection for sparsity basis $\mathbf{\Psi}$ is usually hand-crafted, and the physical parameters of the setup need to be calibrated or chosen very carefully. To resolve these difficulties, we solve Eq. (6) and Eq. (7) using unrolled neuronal networks [32], seeking for a good balance between speed and accuracy. The idea is to turn the fixed, human-tuned algorithmic parameters, as well as the physical parameters of the imaging setup into differentiable system parameters that are automatically tuned via back-propagation through supervised learning.

We now discuss specific solutions (termed network modules) to these two steps, as illustrated in Fig. 2. The input and output of the network are a hologram and the corresponding 3D volume, respectively. Right after the input is a back-propagation initialization block for speeding up network convergence, since compared to random initialization, the back-propagated reconstruction is much more physically

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

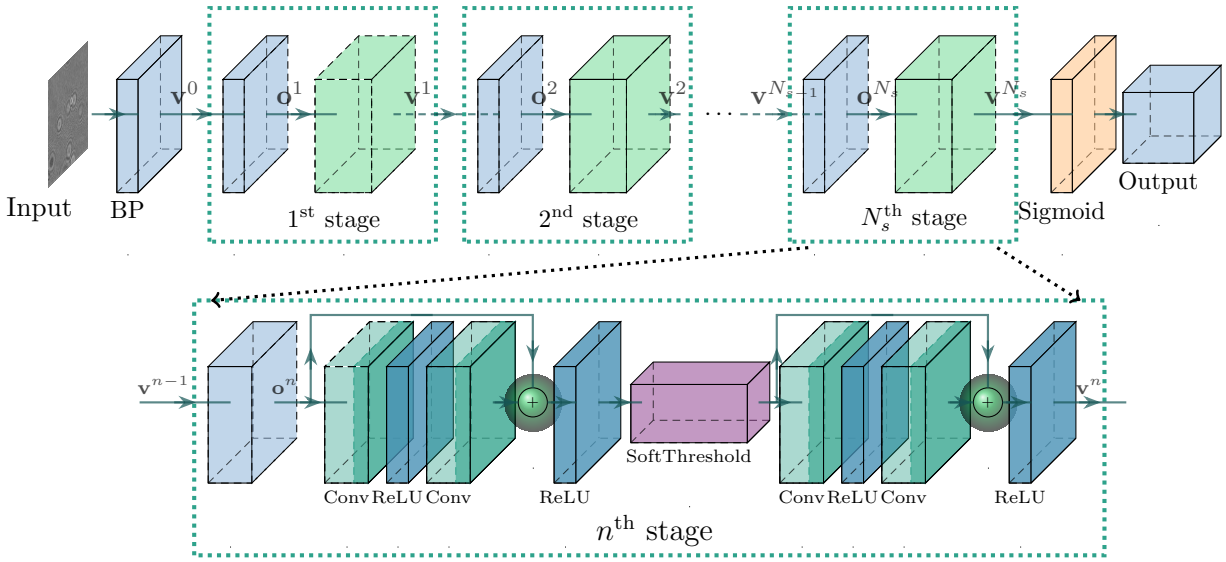


Fig. 2: Framework of the MB-HoloNet with N_s stages. BP: back-propagation.

close to the target volume, though accompanied with some unwanted noise. The network consists of several stages. Each stage consists of two blocks, corresponding to the $\mathbf{o}^n$-module and $\mathbf{v}^n$-module respectively, as presented in the following paragraphs. To increase network flexibility, we allow both γ and μ to vary across iterations, with subscript n denotes for the n^{th} iteration, i.e. the n^{th} stage in the network.

$\mathbf{o}^n$-module corresponds to Eq. (6), which produces an immediate reconstruction $\mathbf{o}^{n+1}$ from the previously calculated $\mathbf{v}^n$. The solution to Eq. (6) is simply a least squares estimate:

$$\mathbf{o}^n = \left(\mathbf{\Phi}^\intercal\mathbf{\Phi} + \mu_n\mathbb{I}\right)^{-1}\left(\mathbf{v}^n + \mu_n\mathbf{\Phi}^\intercal\mathbf{I}_h\right), \tag{8}$$

with $\mathbb{I}$ being the identity matrix. Fortunately, under the assumption of u_r being a plane-wave illumination, the linear operator $\mathbf{P}$ can be rephrased as a convolution followed by sum. We can further break down the above separately for each 2D plane at specific z as:

$$\mathbf{o}^n = [\mathbf{o}^n_{z_1}, \cdots, \mathbf{o}^n_{z_2}]^\intercal, \tag{9}$$

$$\mathbf{o}^n_z = \mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}(\mathbf{p}_z)}\mathcal{F}(\mathbf{I}_h) + \mu_n\mathcal{F}(\mathbf{v}^n_z)}{\mu_n|\mathcal{F}(\mathbf{p}_z)|^2 + \mathbf{1}}\right), \tag{10}$$

where $\mathbf{p}_z$ is the propagation kernel at depth z, and $\mathcal{F}/\mathcal{F}^{-1}$ denote for the forward/inverse Fourier transforms. We initialized $\mathbf{v}^0$ with classical back-propagated solution, i.e. $\mathbf{v}^0 = \mathbf{\Phi}^\intercal\mathbf{I}_h$.

$\mathbf{v}^n$-module aims to compute $\mathbf{v}^n$ from $\mathbf{o}^n$ according to Eq. (7), which is a special case of the so-called proximal mapping $\text{prox}_{\tau\phi}(\cdot)$ where $\phi(\mathbf{v}) = \|\mathbf{\Psi}\mathbf{v}\|_1$. We now consider a general nonlinear transform $\mathcal{G}(\cdot)$ that sparsifies the object. Replacing $\mathbf{\Psi}$ with $\mathcal{G}(\cdot)$, we can obtain a sparsity-inducing regularization problem. Inspired by the powerful representation of convolutional neural networks (CNN) and its universal approximation property [33], we design $\mathcal{G}(\cdot)$ as a block of ResNet, i.e., two 3×3 convolutional layers with the same number of output channels , while the output channel equals to the volume depth, this is similar to 3D kernels from a mathematical perspective. Each convolutional layer is followed by a batch normalization layer and a rectified linear unit (ReLU) activation function, then add the input directly before the final ReLU activation function [34]. This can be represented as:

$$\mathcal{G}(\mathbf{v}) = \text{ReLU}\Big(\text{Conv}\Big(\text{ReLU}\big(\text{Conv}(\mathbf{v})\big)\Big) + \mathbf{v}\Big). \tag{11}$$

To obtain $\mathbf{v}$ from $\mathcal{G}(\mathbf{v})$, we introduce the inverse transformation function $\mathcal{G}^{-1}(\cdot)$, which can be written in the same structure as $\mathcal{G}(\mathbf{v})$. Incorporating $\mathcal{G}(\cdot)$ and $\mathcal{G}^{-1}(\cdot)$ into Eq. (7), we obtain the optimization $\mathbf{v}^n = \mathcal{G}^{-1}\Big(\text{SoftThreshold}\left(\mathcal{G}(\mathbf{o}^n), \tau\right)\Big)$, as in [23].

C. Network structure and loss design

As stated previously, the parameters to be learned are $\Theta = \{\gamma_n, \mu_n, \tau_n, \mathcal{G}_n, \mathcal{G}_n^{-1}\}_{n=1}^{N_s}$. The whole framework of the MB-HoloNet consists of N_s stages, as shown in Fig. 2, with the input being a single 2D hologram image, and output being the desired 3D particle volume to be restored. As mentioned, $\mathbf{o}^n$ denotes the restored 3D object of the n^{th} stage, and $\mathbf{v}^n$ denotes the auxiliary variable in the n^{th} stage. In each stage, the $\mathbf{o}^n$ is updated according to Eq. (6), the estimated immediate reconstruction $\mathbf{o}^n$ is passed to the CNN for the next stage to compute $\mathbf{v}^n$ according to Eq. (7). The filter number of the CNN equals to the number of depth slices of the 3D object. A sigmoid activation layer is implemented before the output to map the object transfer function to the absorption function.

Loss function: $\mathbf{o}$ is the corresponding true value of the the estimated network output $\mathbf{v}^{N_s}$, MB-HoloNet seeks to optimize the following total loss function

$$\mathcal{L} = \mathcal{L}_{\text{model}} + \tau\mathcal{L}_{\text{residual}} \tag{12}$$

given by the squared ℓ_2-norm of the 3D data and a penalty of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

the residual block, to guarantee both data and prior fidelity:

$$\mathcal{L}_{\text{model}} = \frac{1}{N_b N} \sum_{n=1}^{N_b} \|\mathbf{v}^{N_s} - \mathbf{o}\|_2^2, \tag{13}$$

$$\mathcal{L}_{\text{residual}} = \frac{1}{N_b N N_s} \sum_{n=1}^{N_b} \sum_{m=1}^{N_s} \|\mathcal{G}_m^{-1}\mathcal{G}_m(\mathbf{v}^n) - \mathbf{v}^n\|_2^2, \tag{14}$$

where N_b is the block number, N_s is the stage number, and N is the object size.

In summary, with this designed network, at each stage, based on the previously calculated object, the $\mathbf{o}^n$-module and $\mathbf{v}^n$-module optimize the object with a special back-propagation and the following proximal mapping, and the output is transferred to the next stage for further optimization. This structure ensures all parts of the network are based on the imaging model, thus being more powerful. To tackle the issues of limited computational memory and stagnation in local minima during optimization, every time only a small batch of the entire training data is fed into the network, instead of the entire set of holograms. The network is trained using the Adam optimizer [35], which is a form of gradient descent, where the initial learning rate is set empirically, decaying according to the loss descent rate dynamically as the training progresses. We implement the MB-HoloNet using Tensorflow 1.15 and Keras 2.2.5. All experiments were performed on a workstation with an Intel Core i7-6820 CPU and an NVIDIA GTX1080 GPU.

III. Verification

A. Evaluation metrics

We make use of three quantitative evaluation metrics for the assessment of the network performance. These are pairwise correlation coefficient (PCC), structural similarity index for measuring (SSIM) and mean absolute error (MAE). These are defined as:

$$\text{PCC}(x,y) = \frac{\sum_{n=1}^{N}(x_n - \mu_x)(y_n - \mu_y)}{\sqrt{\sum_{n=1}^{N}(x_n - \mu_x)^2}\sqrt{\sum_{n=1}^{N}(y_n - \mu_y)^2}}, \tag{15}$$

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{16}$$

$$\text{MAE}(x,y) = \frac{1}{N}\sum_{n=1}^{N}|x_n - y_n|, \tag{17}$$

where μ_x and μ_y are means of x and y, σ_x^2 and σ_y^2 are the corresponding variances, and σ_{xy} is the covariance of x and y. $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, where $k_1 = 0.01$, $k_2 = 0.03$, L is the dynamic range of the pixel-values. The first two metrics measure how well the samples are likely to be predicted by the model, with best possible scores 1.0, and smaller values indicating worse performance. MAE is an average measure of the absolute difference between two variables and its best possible score is 0.

B. Numerical verification of particle localization

Particle localization is a special case of the more general particle volume reconstruction when the particle and voxel are of the same size. In this section, we verify the feasibility of the proposed MB-HoloNet and its robustness with single-pixel size particles.

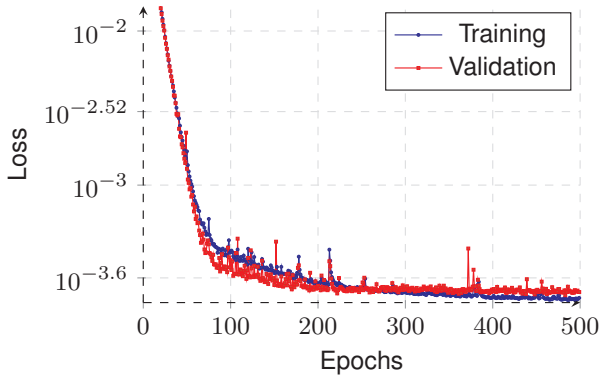


Fig. 3: Training and validation loss decrease along the training process. Only first 500 iterations are shown here.

In the first simulation, one thousand Gabor holograms of randomly distributed particle volumes were synthesized. The lateral and axial resolutions of the particle volumes are $20\,\mu\text{m}$ and $100\,\mu\text{m}$, respectively, and the sensor pixel pitch is the same as volume voxel size. The particle density is 5×10^{-3} particle per voxel (ppv). For each hologram, we have added $50\,\text{dB}$ random Gaussian noise. The generated holograms, along with the 3D particle volumes, were fed into the proposed MB-HoloNet as the training data and label. We split $80\,\%$ as training set and $20\,\%$ for validation. Extra 100 holograms were synthesized with the same specifications for testing. Figure 3 presents one example of the training and validation loss. As seen, the MB-HoloNet converges gradually as the loss decreases along with the training. This agrees with our expectation that the network is continuously updating its parameters and learning representative features of the holograms. We then investigated the affection of stage number N_S. To do that, we performed training with varying stage numbers. The training loss with respect to the stage number are presented in Fig. 4, which demonstrates that convergence depends on the stage number. It converges faster as the stage number increasing. However, when the stage number is larger than 6, the convergence does not improve further. Considering the computational cost, we chose $N_s = 5$ in our following simulations.

Next, We verified the robustness of the MB-HoloNet by showing that the MB-HoloNet works for systems with varying specifications, and the trained network works for holograms captured in different scenarios as the training data. To verify that MB-HoloNet works for varying systems, objects with different dimensions ($32 \times 32 \times 7$, $64 \times 64 \times 32$, and $128 \times 128 \times 30$) were used to synthesize the holograms. In each configuration, the wavelength is fixed to $\lambda = 660\,\text{nm}$. Specifications of the training parameters and the test results are in table table I. Figure 5 shows several selected ground-truths and the corresponding MB-HoloNet predictions with the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

TABLE I: Training details of different datasets.

Cases	$N_x \times N_y \times N_z$	$32 \times 32 \times 7$	$64 \times 64 \times 32$	$128 \times 128 \times 30$
Train	ppv ($\times 10^{-3}$)	$1 \sim 5$	$0.2 \sim 1$	$0.5 \sim 1$
	Data sets	1000	1000	1000
	N_s	5	5	5
	Batch size	32	32	32
	Learning rate ($\times 10^{-3}$)	2	1	2
	τ ($\times 10^{-4}$)	1	1	1
	Epochs	1000	1500	1000
Results	Training time	20 min	3 h30 min	5 h31 min
	Test time	12 ms	17 ms	41 ms
	Test score	99.91	99.67	99.42
	Memory cost (megabytes)	223	4057	15218

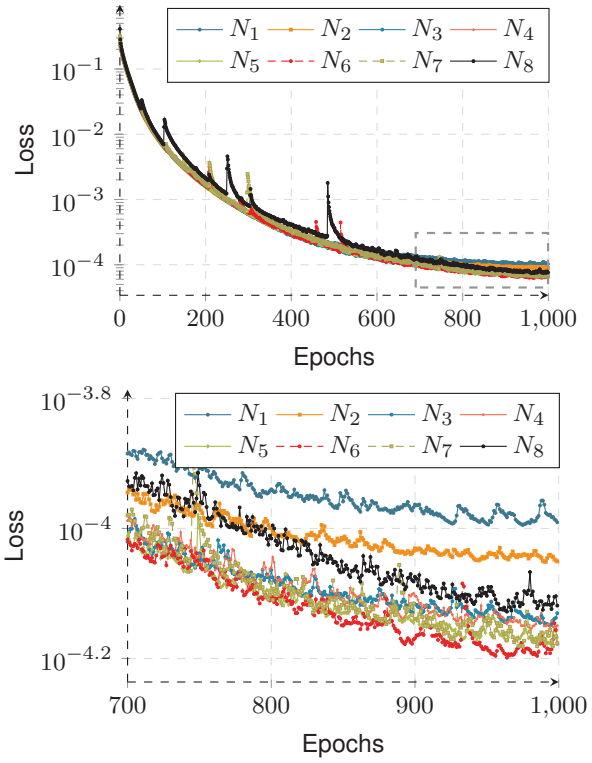
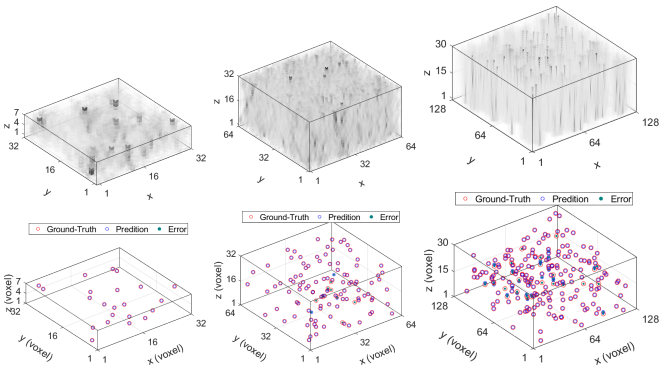


Fig. 4: Loss along the training process with respect to varying stage number N_s. The bottom figure is the part within the gray dashed rectangle in the above figure.



(a) $32 \times 32 \times 7$ (b) $64 \times 64 \times 32$ (c) $128 \times 128 \times 30$

Fig. 5: Hologram localization reconstruction with three different dimensions, synthesized with various specifications. The above images are the back-propagated reconstructions. In the below images, the red, blue circles are the predictions and ground-truth, and the green dots in red and blue circles are the unpaired ground-truth and the false predictions.

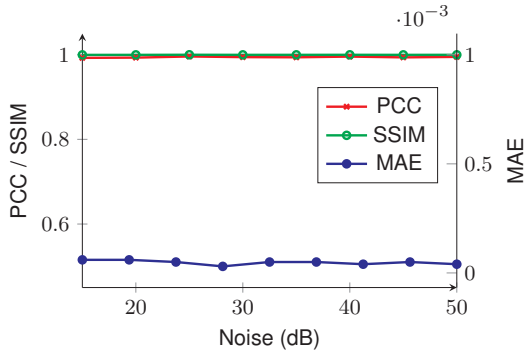


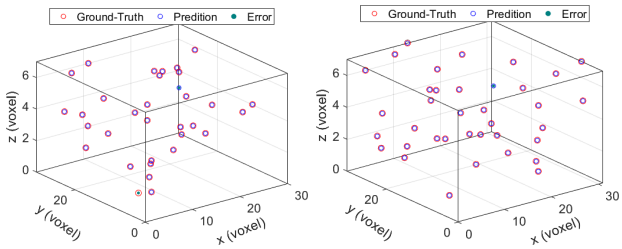
Fig. 6: Test accuracy with respect to varying noise levels.

back-propagated initial volumes, while the first row shows the back-propagation reconstruction and the second row shows the predicted locations compared to the ground-truths. Comparing to ground-truths, we conclude that MB-HoloNet works well for all the present cases. It should be mentioned that the memory usage issue is inevitable in 3D imaging. Representing 3D volume using 2D depth maps may be one solution, but it can not deal with occlusion in a 3D volume. Our approach represents the 3D volume explicitly, and hence the memory usage in the training stage increases with the volume size, as shown in the bottom row of Table I. Yet this memory limitation issue can be resolved by sequential training on multiple GPUs or on CPU clusters.

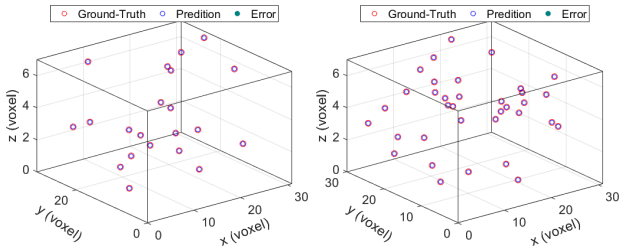
Furthermore, we test the trained network with holograms synthesized by the same system but with different noise levels and different ppvs. We choose the trained model in Fig. 5 (a) for the test. In the first test, seven groups of holograms with added noise from 10 dB to 50 dB were fed to the network. The quality of the network outputs are measured with the metrics presented in Section III-A. Figure 6 plots the accuracy metric concerning varying noise levels, and Fig. 7 shows randomly selected scattering plot of the predictions and ground-truth. In Figs. 6 and 8, the metrics values were averaged over all of the test data. It shows that when the noise SNR is much lower than that in the training data (50 dB), we can still reach a reasonable image quality. This indicates that the network is

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

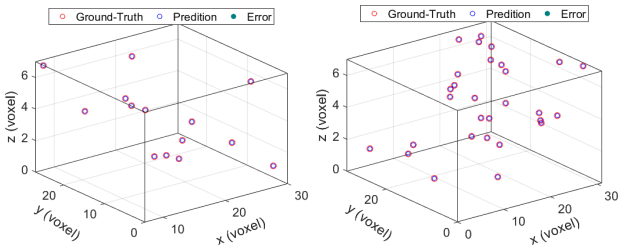


(a) 10 dB (b) 20 dB

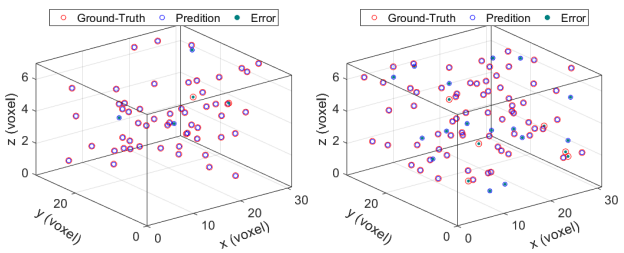


(c) 30 dB (d) 40 dB

Fig. 7: MB-HoloNet particle reconstruction vs. ground-truth under varying noise. The red, blue circles are the predictions and ground-truth, and the green dots in red and blue circles are the unpaired ground-truth and the false predictions.



(a) 2×10^{-3} (b) 4×10^{-3}



(c) 8×10^{-3} (d) 10×10^{-3}

Fig. 9: MB-HoloNet particle reconstruction vs. the ground-truth under varying ppv. The red, blue circles are the predictions and ground-truth, and the green dots in red and blue circles are the unpaired ground-truth and the false predictions.

numerically stable to noise, at ppv of $1 \sim 5 \times 10^{-3}$. Given that the reconstruction volume consists of sparse particles, in this situation, all metrics are consistent with respect to the noise level, indicating that our proposed approach is robust to noise for this particular ppv.

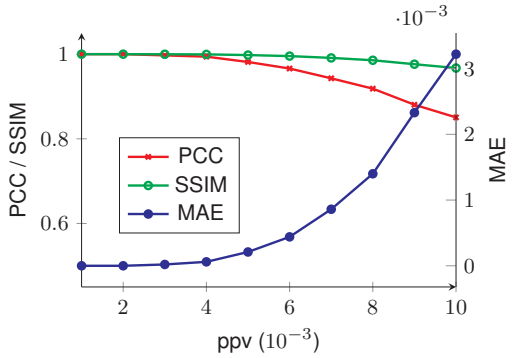
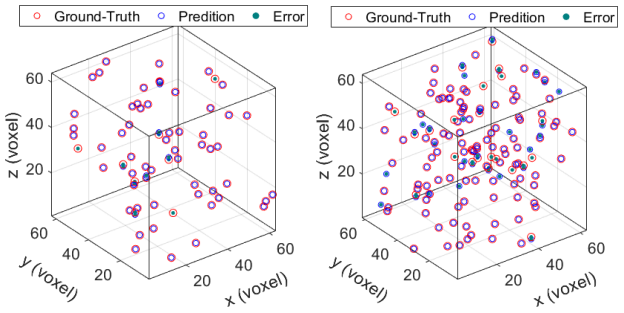


Fig. 8: Test accuracy with respect to varying ppv.

Figure 8 plots the accuracy metrics with respect to varying ppv and Fig. 9 shows several selected scatter plots of the predictions and ground-truth. They show that when the ppv is lower than in the training data, the trained network can always reconstruct the particles. However, the image quality degrades while the ppv increasing in the testing data, while the SSIM can still maintain a high degree even if the particle density of the test data is twice that of the training data.

One more simulation was performed for comparing the particle localization with the state-of-the-art work [17]. In the reference paper, volumes with particle per pixel (ppp) of $1.9 \times 10^{-4} \sim 6.1 \times 10^{-2}$ were tested. This corresponds to a ppv of $1.5 \times 10^{-6} \sim 4.8 \times 10^{-4}$. We have synthesized holograms for particles with the same ppv and dimension of



(a) ppp $= 1.6 \times 10^{-2}$ (b) ppp $= 6.1 \times 10^{-2}$

Fig. 10: Comparison of prediction results with different ppp's. The red, blue circles are the predictions and ground-truth, and the green dots in red and blue circles are the unpaired ground-truth and the false predictions.

$64 \times 64 \times 64$, trained the network with 1500 data-sets for 2000 epochs. Figure 10 presents the comparison between the ground-truths and the network predictions, while Fig. 10(a) and (b) are with ppp of 1.6×10^{-2} and 6.1×10^{-2} respectively, the location extraction rates are 87.5% and 82.4%, which are lower than the 95% in the reference paper.

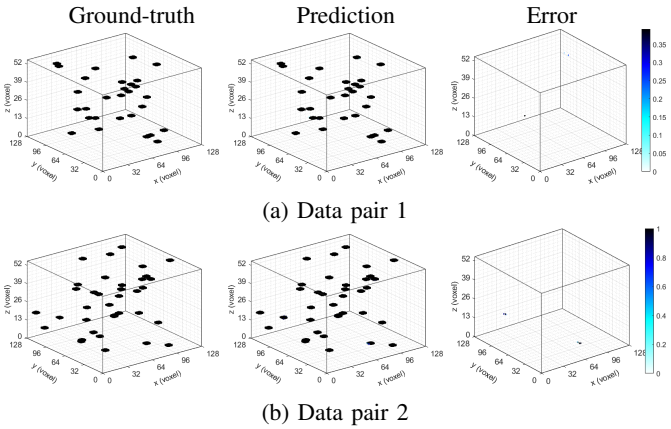
C. Numerical verification of particle size reconstruction

We verify the particle reconstruction by setting the particle size larger than a single voxel. In preparing the training data, the particles were assumed as disks with a radius of $50\,\mu\text{m}$, with a camera pixel pitch of $7\,\mu\text{m}$. Therefore, the reconstructed particles should be disks and their radius should cover about seven image pixels. The training process is the same as described in previous sections. We have tested the trained network after 1500 epochs. Two randomly selected ground-truth and the predictions are shown in Fig. 11. The

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

reconstructions indicate that the circular shape particles, and the pixel count of the disks' radius are approximately seven, as expected, revealing a successful reconstruction of the particle shape and size. From the error plots, we can see that the particle size have been well reconstructed.



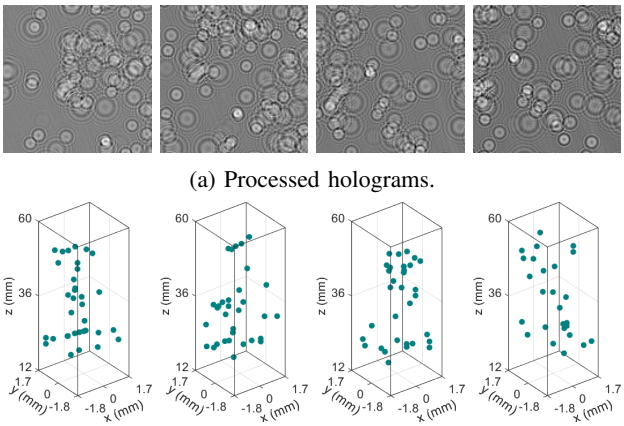
(a) Data pair 1

(b) Data pair 2

Fig. 11: Comparison of particle size reconstruction.

IV. Experimental verification

A. Particle localization



(a) Processed holograms.

(b) Locations extracted by the MB-HoloNet.

Fig. 12: Experimental holograms (a), Back-propagation reconstructions (a), and the corresponding extracted particle locations with the MB-HoloNet (b). Refer to Visualization 1, 2, 3, 4 for a complete 3D visualization.

In the first experiment, particles with a diameter of $50\,\mu\text{m}$ were randomly seeded in water. The particles were located at a distance ranging from $12\,\text{mm}$ to $60\,\text{mm}$ to the image sensor. A laser with a wavelength of $660\,\text{nm}$ illuminated the particles and form the hologram captured by the CCD, as shown in Fig. 12(a). The pixel pitch and resolution of the CCD are $3.45\,\mu\text{m}$ and 1024×1024 respectively. The captured hologram is subtracted by the background and scaled using bicubic interpolation to a size of 128×128 pixels. Since the particles are very sparse, the depth interval of the view interest was set to $1\,\text{mm}$, which corresponding to a volume of $128 \times 128 \times 49$ voxels. We generated one thousand and five hundred holograms with the same specifications as in the processed hologram and trained the network with the synthetic holograms. After 1000 epochs training (took 12 hours), we test the MB-HoloNet with the experimental captured hologram. Figure 12 shows the results, while Fig. 12(a) shows four processed holograms and Fig. 12(b) shows the reconstructed particle locations.

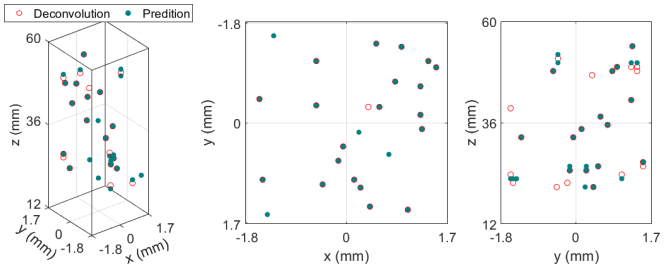


Fig. 13: Location comparison of reconstructed particles between FASTA deconvolution and MB-HoloNet.

Since there is no ground-truth for the real particles, we verify the predicted particle locations by comparing them to the reconstructions of the previously developed methods, which are also not fully accurate but only for empirical comparison. Comparing to the extended-focused-imaging [36]–[38], the deconvolution approaches are more accurate and practical for 3D particle imaging [5], [8]. Here we choose the state-of-the-art FASTA deconvolution method [8]. Figure 12 shows the results after FASTA performed 1000 iterations. The left image shows the 3D view of the location comparison, and the middle and right images show the top and side view of the left image, respectively. The red circles and green dots represent the FASTA deconvolution and MB-HoloNet, respectively. We can see that most of the particle locations predicted by the MB-HoloNet match the FASTA deconvolution results. The axial locations are not as good as the lateral ones, but this is reasonable because the axial resolution is always worse than the lateral resolution [27]. Given similar reconstruction quality, the MB-HoloNet only took $85\,\text{ms}$ while the FASTA deconvolution took $173\,\text{s}$ for 1000 iterations.

B. Particle size reconstruction

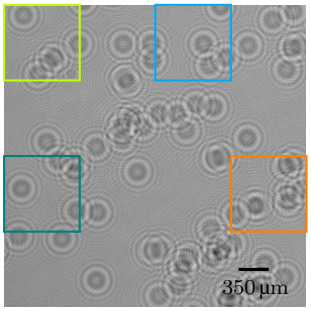
In the second experiment, we show that the proposed MB-HoloNet can reconstruct particle size and locations simultaneously. The particles with diameters ranging from $45\,\mu\text{m}$ to $53\,\mu\text{m}$ were adhered on a microscopy slide. The slide was put at a distance of $19.7\,\text{mm}$ from the CCD camera. The pixel pitch and resolution of the CCD are $3.45\,\mu\text{m}$ and 1024×1024 respectively. The holograms were cropped to 256×256 pixels and then resized to 128×128 pixels. Thus the diameter of the reconstructed particles from the processed holograms should be $7\sim 8$ pixels. For the training data, volumes span a depth of $14\,\text{mm}$ to $25\,\text{mm}$ were chosen to synthesize the holograms. All other parameters matched the processed hologram. A total of 1500 data-sets were used to train the network for 1000 epochs, which took $4\,\text{h}\ 32\,\text{min}$.

Figure 14 shows the results. Fig. 14(a) is the original hologram, Fig. 14(b) shows several sub holograms cropped

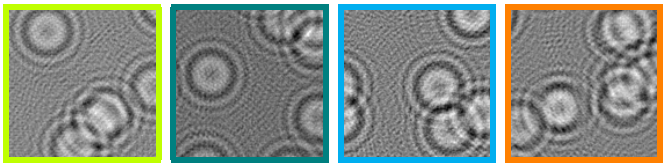
This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

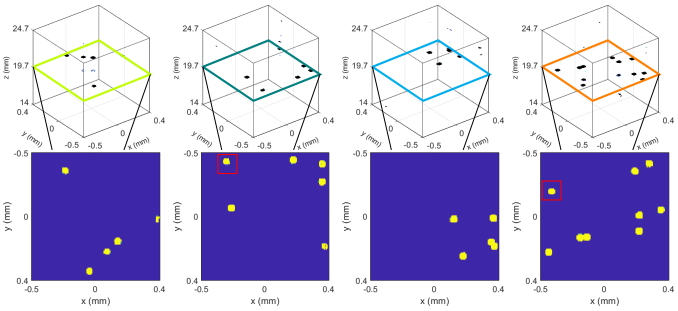
from the original one, and Fig. 14(c) is the corresponding network predictions. The bottom row of Fig. 14(c) is the corresponding slice views of the above row. From the above row, we can see that the particles were reconstructed at the correct locations ($z = 19.7\,\mathrm{mm}$). Compare the bottom row with the corresponding holograms in Fig. 14(b), we see that most of the particles were reconstructed, the red rectangles shows the error predicated particles, while the error amount is at a low rate. The lateral pixels of each particle in Fig. 14(c) is seven, which matches the expectation.



(a) Processed original hologram.



(b) Cropped and resized sub-holograms.



(c) MB-HoloNet predictions.

Fig. 14: Captured hologram (a), processed holograms (b), hologram reconstruction with MB-HoloNet (c).

V. Discussion and Conclusions

We have presented MB-HoloNet for 3D particle imaging with Gabor holography. The MB-HoloNet takes a single 2D hologram as input and outputs the corresponding 3D particle volume. Compared to state-of-the-art technique [17], our proposed MB-HoloNet involves the prior of the underlying imaging model in network design and is able to reconstruct particle size and perform particle localization at the same time, even though the location extraction rate is slightly lower than the reference work. The physical prior employed makes it possible to train the network with a relatively small amount of training data (1000 $\sim$ 1500 in all of the presented cases in this paper). Besides, the single 2D image input and 3D volume output is simple with no time-consuming pre- or post-processing required. Most importantly, the model trained with synthesized data can process experimentally captured holograms; this is superior to previously published works.

While the proposed MB-HoloNet has superior performance compared to the state-of-the-art 3D hologram reconstruction methods, there remains room for further improvements: (1) For the MB-HoloNet, training labels are 3D, which is GPU memory consuming and limits the total volume size that can be processed. The sparsity of the particles may be considered to alleviate the memory consumption, promote training speed, and more massive volume processing. (2) Due to the available small volumes mentioned in (1), the required input holograms have to be small. For this purpose, captured holograms (usually of a larger size) have to be cropped or resized. The cropped holograms cover a small imaging region of a few particles, also the resizing decrease hologram details, and usually leads to induced resolution. It is our next work to find an alternative 3D representation to overcome these limitations.

Acknowledgement

The authors would like to thank Dr. Yuqi Li for his contributions to an earlier version of loop-unrolled networks for digital holography, which has inspired this work [39].

References

[1] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, no. 4098, pp. 777–778, may 1948.
[2] E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory," *Journal of the Optical Society of America*, vol. 52, no. 10, p. 1123, oct 1962.
[3] I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," *Optics Letters*, vol. 22, no. 16, p. 1268, aug 1997.
[4] T.-C. Poon, M. H. Wu, K. Shinoda, and Y. Suzuki, "Optical scanning holography," *Proceedings of the IEEE*, vol. 84, no. 5, pp. 753–764, may 1996.
[5] D. J. Brady, K. Choi, D. L. Marks, R. Horisaki, and S. Lim, "Compressive holography," *Optics Express*, vol. 17, no. 15, pp. 13 040–13 049, Jul 2009.
[6] E. Y. Lam, X. Zhang, H. Vo, T.-C. Poon, and G. Indebetouw, "Three-dimensional microscopy and sectional image reconstruction using optical scanning holography," *Applied Optics*, vol. 48, no. 34, p. H113, oct 2009.
[7] N. Chen, E. Y. Lam, T.-C. Poon, and B. Lee, "Sectional hologram reconstruction through complex deconvolution," *Optics and Lasers in Engineering*, vol. 127, p. 105945, apr 2020.
[8] K. Mallery and J. Hong, "Regularized inverse holographic volume reconstruction for 3d particle tracking," *Optics Express*, vol. 27, no. 13, p. 18069, jun 2019.
[9] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, p. 921, jul 2019.
[10] T. C. Nguyen, V. Bui, and G. Nehmetallah, "3d optical diffraction tomography using deep learning," *Optical Engineering*, vol. 57, no. 4, p. 043111, 2018.
[11] M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," *Scientific Reports*, vol. 7, no. 1, dec 2017.
[12] Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica*, vol. 5, no. 4, p. 337, mar 2018.
[13] R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Optics Express*, vol. 24, no. 13, p. 13738, jun 2016.
[14] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Science & Applications*, vol. 8, no. 1, p. 23, 2019.
[15] Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning in holography and coherent imaging," *Light: Science & Applications*, vol. 8, no. 1, sep 2019.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCI.2021.3063870, IEEE Transactions on Computational Imaging

[16] Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," *Optica*, vol. 4, no. 11, p. 1437, nov 2017.

[17] S. Shao, K. Mallery, S. S. Kumar, and J. Hong, "Machine learning holography for 3D particle field imaging," *Optics Express*, vol. 28, no. 3, pp. 2987–2999, 2020.

[18] S. Shao, K. Mallery, and J. Hong, "Machine learning holography for measuring 3d particle distribution," *Chemical Engineering Science*, vol. 225, p. 115830, nov 2020.

[19] T. Shimobaba, T. Takahashi, Y. Yamamoto, Y. Endo, A. Shiraki, T. Nishitsuji, N. Hoshikawa, T. Kakue, and T. Ito, "Digital holographic particle volume reconstruction using a deep neural network," *Applied Optics*, vol. 58, no. 8, p. 1900, mar 2019.

[20] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1333–1347, jun 2018.

[21] J. Sun, H. Li, Z. Xu *et al.*, "Deep admm-net for compressive sensing mri," in *Advances in neural information processing systems*, 2016, pp. 10–18.

[22] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein, "Unrolled optimization with deep priors," *arXiv preprint arXiv:1705.08041*, 2017.

[23] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1828–1837.

[24] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, nov 2017.

[25] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Optics Express*, vol. 27, no. 20, p. 28075, sep 2019.

[26] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[27] J. W. Goodman, *Introduction to Fourier Optics*. W.H.Freeman & Co Ltd, 2017.

[28] L. Denis, D. Lorenz, E. Thiébaut, C. Fournier, and D. Trede, "In-line hologram reconstruction with sparsity constraints," *Optics Letters*, vol. 34, no. 22, p. 3475, nov 2009.

[29] Y. Endo, T. Shimobaba, T. Kakue, and T. Ito, "GPU-accelerated compressive holography," *Optics Express*, vol. 24, no. 8, p. 8437, apr 2016.

[30] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.

[31] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[32] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th international conference on international conference on machine learning*, 2010, pp. 399–406.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, may 2017.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, jun 2016, pp. 770–778.

[35] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. The MIT Press, 2017.

[36] W. Chen, C. Quan, and C. J. Tay, "Extended depth of focus in a particle field measurement using a single-shot digital hologram," *Applied Physics Letters*, vol. 95, no. 20, p. 201103, nov 2009.

[37] Y. Wu, X. Wu, J. Yang, Z. Wang, X. Gao, B. Zhou, L. Chen, K. Qiu, and K. Cen, "Wavelet-based depth-of-field extension, accurate autofocusing, and particle pairing for digital inline particle holography," *Applied Optics*, vol. 53, no. 4, p. 556, jan 2014.

[38] J. Lu, J. P. Fugal, H. Nordsiek, E. W. Saw, R. A. Shaw, and W. Yang, "Lagrangian particle tracking in three dimensions via single-camera in-line digital holography," *New Journal of Physics*, vol. 10, no. 12, p. 125013, dec 2008.

[39] N. Chen, Y. Li, and W. Heidrich, "Physics-based holo-net for three-dimensional imaging," in *Imaging and Applied Optics Congress*. Optical Society of America, 2020, p. JTh3D.3. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=DH-2020-JTh3D.3

Ni Chen received the B.S. degree in software engineering from Harbin Institute of Technology, China, in 2008, the M.S. and Ph.D. degrees in electrical engineering from Chungbuk National University and Seoul National University, Korea, in 2010 and 2014, respectively.

She is currently a Research Fellow at the King Abdullah University of Science and Technology. From 2014 to 2016, she was a Research Scientist with the University of Hong Kong, and from 2016 to 2017, and she was an Associate Professor with the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. From 2018 to 2019, she was a Research Assistant Professor in the Department of Electrical and Computer Engineering at Seoul National University, Seoul, Korea. Her research interests include three-dimensional optical imaging and display.

Congli Wang received the B.Eng. degree in electrical engineering from Tianjin University in 2015 and the M.Sc. in electrical engineering from King Abdullah University of Science and Technology in 2016. He is currently working towards his Ph.D. degree under the supervision of Dr. W. Heidrich. His research interests include computational imaging, wavefront sensing, and adaptive optics.

Wolfgang Heidrich is a Professor of Computer Science and the Director of the Visual Computing Center at King Abdullah University of Science and Technology (KAUST), as well as the interim leader for KAUST's AI Initiative. Dr. Heidrich joined KAUST in 2014, after 13 years as a faculty member at the University of British Columbia. He received his Ph.D. from the University of Erlangen in 1999, and then worked as a Research Associate in the Computer Graphics Group of the Max-Planck-Institute for Computer Science in Saarbrucken, Germany, before joining UBC in 2000. Dr. Heidrich's research interests lie at the intersection of imaging, optics, computer vision, computer graphics, and inverse problems. His more recent interest is in computational imaging, focusing on hardware-software co-design of the next generation of imaging systems, with applications such as High-Dynamic Range imaging, compact computational cameras, hyperspectral cameras, to name just a few. Dr. Heidrich's work on High-Dynamic Range Displays served as the basis for the technology behind Brightside Technologies, which was acquired by Dolby in 2007. Dr. Heidrich has chaired the papers program for both Siggraph Asia and the International Conference of Computational Photography (ICCP) among others. He is the recipient of a 2014 Humboldt Research Award.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/