

GPU-based deep convolutional neural network for tomographic phase microscopy with ℓ_1 fitting and regularization

Hui Qiao
Jiamin Wu
Xiaoxu Li
Morteza H. Shoreh
Jingtao Fan
Qionghai Dai

GPU-based deep convolutional neural network for tomographic phase microscopy with ℓ_1 fitting and regularization

Hui Qiao,^a Jiamin Wu,^a Xiaoxu Li,^a Morteza H. Shoreh,^b Jingtao Fan,^a and Qionghai Dai^{a,*}

^aTsinghua University, Department of Automation, Beijing, China

^bÉcole Polytechnique Fédérale de Lausanne, School of Engineering, Laboratory of Optics, Lausanne, Switzerland

Abstract. Tomographic phase microscopy (TPM) is a unique imaging modality to measure the three-dimensional refractive index distribution of transparent and semitransparent samples. However, the requirement of the dense sampling in a large range of incident angles restricts its temporal resolution and prevents its application in dynamic scenes. Here, we propose a graphics processing unit-based implementation of a deep convolutional neural network to improve the performance of phase tomography, especially with much fewer incident angles. As a loss function for the regularized TPM, the ℓ_1 -norm sparsity constraint is introduced for both data-fidelity term and gradient-domain regularizer in the multislice beam propagation model. We compare our method with several state-of-the-art algorithms and obtain at least 14 dB improvement in signal-to-noise ratio. Experimental results on HeLa cells are also shown with different levels of data reduction. © The Authors.

Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.23.6.066003](https://doi.org/10.1117/1.JBO.23.6.066003)]

Keywords: tomographic phase microscopy; ℓ_1 data-fidelity; GPU-based implementation; neural network; refractive index.

Paper 180094R received Feb. 13, 2018; accepted for publication May 21, 2018; published online Jun. 14, 2018.

1 Introduction

Most biological samples such as live cells have low contrast in intensity but exhibit strong phase contrast. Phase contrast microscopy is then widely applied in various biomedical imaging applications.¹ In the past decades, the development of quantitative phase imaging^{2,3} gives rise to a label-free imaging modality, tomographic phase microscopy (TPM), which deals with the three-dimensional (3-D) refractive index distribution of the sample.^{4–6} The label-free and noninvasive character makes it attractive in biomedical imaging, especially for cultured cells.^{7,8}

However, most of the current methods require around 50 quantitative phase images acquired at different angles^{9–11} or different depths⁶ for optical tomography. This speed limitation greatly restricts its field of applications. For example, the difference of the refractive index may be blurred during the angular (or axial) scanning when observing fast-evolving cell dynamics or implementing high-throughput imaging cytometry.¹¹ Another challenge for TPM is the missing cone problem, which limits its reconstruction performance, especially for limited axial resolution compared with the subnanometer optical-path-length sensitivity.¹²

To relieve the missing cone problem, many methods have been developed for better signal-to-noise ratio (SNR) with fewer images. Different regularizations such as the positivity of the refractive index differences^{4,13} and the sparsity in some transform domain^{14,15} are added to an iterative reconstruction framework based on the theory of diffraction tomography,^{16,17} for reducing the artifacts induced by the missing cone problem

and the limited sampling rates in Fourier domain. Both intensity-coded and phase-coded structured illumination methods further promote the performance by their better multiplexing ability compared with conventional plane-wave illumination.^{18,19} However, these methods suffer from the great degradation when the scattering effects become significant in the sample. The beam propagation method (BPM)²⁰ is then applied in phase tomography to provide a more accurate model by considering the nonlinear light propagation with scattering.^{21,22} And the multislice propagation modeling is definitely similar to the neural network in the field of machine learning.^{23,24} By combining the nonlinear modeling and the sparse constraint in the gradient domain, the Psaltis group has validated the competitive capability of this learning approach over conventional methods.^{21,23} Despite its success in modeling with ℓ_2 -norm constraint, the current method is still a preliminary network, especially compared with the state-of-the-art deep learning frameworks,²⁵ and the iterative reconstruction is challenging to deploy in practice due to the high computational cost and the difficulty of the hyperparameter selection. More potential can be exploited in both optimization algorithms and better network architectures.

In this paper, we propose a graphics processing unit (GPU)-based implementation of a deep convolutional neural network (CNN) to simulate the multislice beam propagation for TPM. A loss function consisting of an ℓ_1 -norm data-fidelity term and an ℓ_1 -norm gradient-domain regularizer is devised to achieve higher reconstruction quality even with fewer training data. To deal with the vast quantities of parameters and regularizers, we apply the adaptive moment estimation (Adam) algorithm²⁶ for optimization, which can also be regarded as the training process of the CNN. Compared with previous works using stochastic gradient descent,^{23,24} our method ensures

*Address all correspondence to: Qionghai Dai, E-mail: qhddai@tsinghua.edu.cn

a faster convergence and a better robustness to the initial value. Both simulation and experimental results on polystyrene beads, and HeLa cells are shown to validate its reconstruction performance. We anticipate that our work can not only boost the performance of optical tomography, but also guide more applications of deep learning in the optics field.

2 Materials and Methods

2.1 Experimental Setup

Figure 1 shows the schematic diagram of the experimental setup. In our system,²³ the sample placed between two cover glasses is illuminated sequentially at multiple angles and the scattered light is holographically recorded. A laser beam ($\lambda = 561$ nm) is split into sample and reference arms by the first beam splitter. In the sample arm, a galvo mirror varies the angle of illumination on the sample using the 4F system created by L1 and OB1. The light transmitted through the sample is imaged onto the CMOS camera via the 4F system created by OB2 and L2. The beam splitter (BS2) recombines the sample and reference laser beams, forming a hologram at the image plane. The numerical apertures (NAs) of OB1 and OB2 are 1.45 and 1.4, respectively. For data acquisition, we capture multiple tomographic phase images by near-plane-wave illumination (Gaussian beam) with equally spaced incident angles. We use a differential measurement between the phase on a portion of the field of view on the detector that does not include the cell and the cell itself to maintain phase stability. Accordingly, complex amplitudes extracted from the measurements constitute the training set of our proposed CNN.

2.2 Beam Propagation Method

We build the CNN, based on the forward model of light propagation,^{21,23} to model the diffraction and propagation effects of light-waves. It is known that the scalar inhomogeneous Helmholtz equation completely characterizes the light field at all spatial positions in a time-independent form

$$[\nabla^2 + k^2(\mathbf{r})]u(\mathbf{r}) = 0, \quad (1)$$

where $\mathbf{r} = (x, y, z)$ denotes a spatial position, u is the total light-field at \mathbf{r} , $\nabla^2 = (\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2})$ is the Laplacian, and $k(\mathbf{r})$ is the wave number of the light field at \mathbf{r} . The wave number depends on the local refractive index distribution $n(\mathbf{r})$ as

$$k(\mathbf{r}) = k_0 n(\mathbf{r}) = k_0 [n_0 + \delta n(\mathbf{r})], \quad (2)$$

where $k_0 = 2\pi/\lambda$ is the wave number in vacuum, n_0 is the refractive index of the medium, and the local variation $\delta n(\mathbf{r})$ is caused by the sample inhomogeneities. By introducing the complex envelope $a(\mathbf{r})$ of the paraxial wave $u(\mathbf{r}) = a(\mathbf{r}) \exp(jk_0 n_0 z)$ for BPM, we can obtain an evolution equation²¹ in which z plays the role of evolution parameter

$$a(x, y, z + \delta z) = e^{jk_0 \delta n(\mathbf{r}) \delta z} \times \left[\mathcal{F}^{-1} \left\{ e^{-j \left(\frac{\omega_x^2 + \omega_y^2}{k_0 n_0 + \sqrt{k_0^2 n_0^2 - \omega_x^2 - \omega_y^2}} \right) \delta z} \right\} * a(\cdot, \cdot, z) \right], \quad (3)$$

where δz is a sufficiently small but a finite z step, ω_x and ω_y represent angular frequency coordinates in the Fourier domain, $a(\cdot, \cdot, z)$ expresses the two-dimensional (2-D) complex envelope at z depth, $*$ refers to a convolution operator, and $\mathcal{F}^{-1}\{\cdot\}$ means the 2-D inverse Fourier transform.

2.3 GPU-Based Implementation of CNN

A schematic architecture of our CNN is shown in Fig. 2. For constructing our neural network, we divide the computational sample space into thin slices with the sampling interval δz along the propagation direction z . One slice corresponds to one layer in CNN. Within each layer, neurons specify the discretized light-field with transverse sampling intervals δx and δy , respectively. The input layer is the incident field upon the sample. In terms of the Eq. (3), inputs are then passed from nodes of each layer to the next, with adjacent layers connected by alternating operations of convolution and multiplication. At the very last layer of our CNN, the output complex field amplitude is then bandlimited by the NA of the imaging system composed

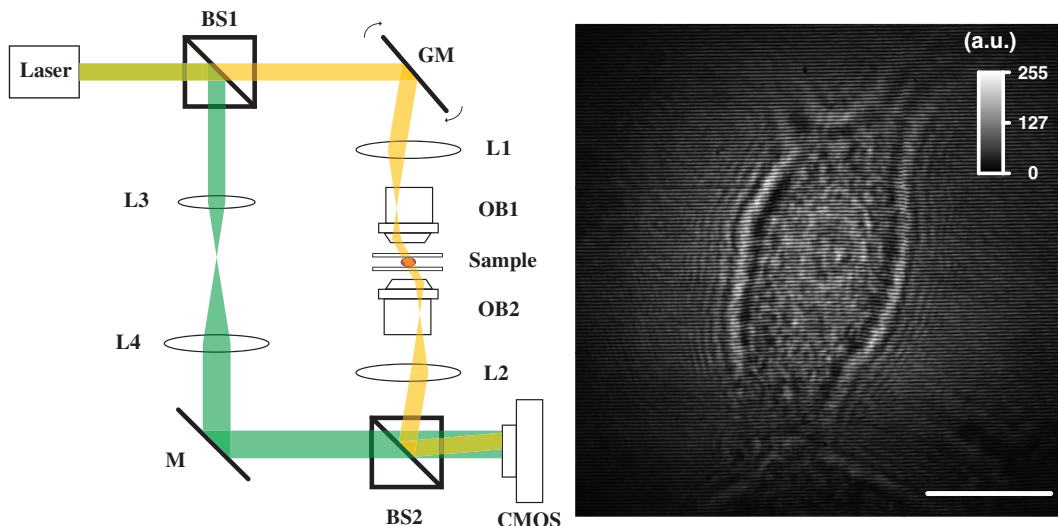


Fig. 1 Experimental setup (BS, beam splitter; GM, galva mirror; L, lens; M, mirror; and OB, objective) and measured hologram by CMOS. Scale bar, 10 μm .

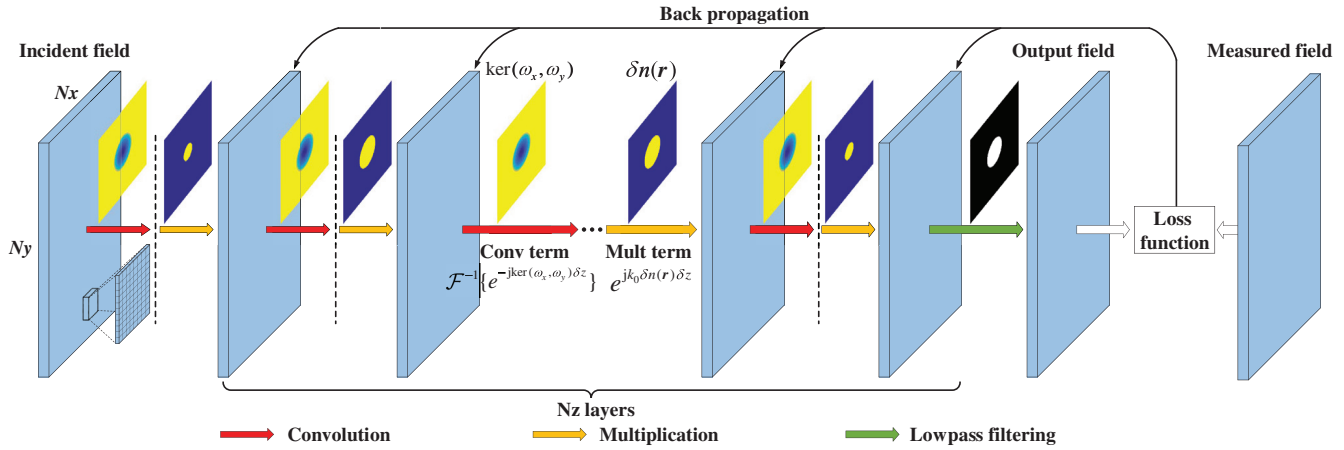


Fig. 2 Detailed schematic of our CNN architecture, indicating the number of layers (N_z), nodes ($N_x \times N_y$) in each layer and operations between adjacent layers. Here, $\ker(\omega_x, \omega_y)$ signifies $(\omega_x^2 + \omega_y^2)/(k_0 n_0 + \sqrt{k_0^2 n_0^2 - \omega_x^2 - \omega_y^2})$ and we take the $\delta n(\mathbf{r})$ of a polystyrene bead as example.

of lenses OB2 and L2 in Fig. 1. We implement the proposed network on the basis of TensorFlow framework. The connection weight $\delta n(\mathbf{r})$ can be trained using the Adam algorithm for optimization on the following minimization problem:

$$\min_{\delta n} \frac{1}{M} \sum_{m=1}^M \|Y_m(\delta n) - G_m(\delta n)\|_1 + \tau R(\delta n)$$

$$\text{s.t. } R(\delta n) = \sum_r \|\nabla \delta n(\mathbf{r})\|_1 \quad \text{and} \quad \delta n \geq 0, \quad (4)$$

where M denotes the number of measured views, $\|\cdot\|_1$ indicates the ℓ_1 -norm, and $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ is the differential operator. For a given view m , Y_m and G_m are the output of the last layer and the actual measurement acquired by the optical system, respectively. The design of our loss function will be specifically discussed in Sec. 4.1. Compared with the ℓ_2 -norm, the ℓ_1 data-fidelity term relaxes the intrinsic assumptions on the distribution of noise (symmetry and no heavy tails) and suits better for the measurements containing outliers. Hence, it can be effectively applied to the biomedical imaging especially when the noise model is heavy-tailed and undetermined.²⁷ As a regularization term, $R(\delta n)$ imposes the ℓ_1 -norm sparsity constraint on a gradient domain according to its better characteristic for the reconstruction from higher incomplete frequency information than ℓ_2 -norm,^{28,29} whereas τ is the positive parameter controlling the influence of regularization. The positivity constraint takes advantage of the assumption that the index perturbation is real and positive when imaging weakly absorbing samples such as biological cells. The subgradient method³⁰ plays an important role in machine learning for solving the optimization framework under ℓ_1 -norm and the Ref. 26 has verified the theoretical convergence properties of the Adam algorithm, which will be specifically discussed in Sec. 4.3. We perform the neural network computations on 4 NVIDIA TITAN Xp graphics cards and the processing time to run the learning algorithm (100 iterations) on $256 \times 256 \times 160$ nodes is nearly 9 min. Obviously, it is possible to make the optimization of hyperparameters, which have an important effect on results, a more reproducible and automated process and thus is beneficial for training the large-scale and often deep multilayer neural networks

successfully and efficiently.³¹ The full implementation and the trained networks are available at <https://github.com/HuiQiaoLightning/CNNforTPM>.

3 Results

For demonstration, we evaluate the designed network by both simulation and experimental results of the TPM as described before. To make a reasonable comparison, selected hyperparameters have been declared for all the other reconstruction methods. The selection of hyperparameters will be specifically discussed in Sec. 4.2.

3.1 Tomographic Reconstruction of Simulated Data

In simulation, we consider a situation of three $5 \mu\text{m}$ beads of refractive index $n = 1.548$ immersed into oil of refractive index $n_0 = 1.518$ shown in Fig. 3. The centers of the beads are placed at $(0, 0, -3)$, $(0, 0, 3)$, and $(0, 5, 0)$, respectively, with the unit of micron. The training set of the framework is simulated as 81 complex amplitudes extracted from the digital-holography measurements with different angles of incidence evenly distributed in $[-\pi/4, \pi/4]$ by BPM, whereas the illumination is tilted perpendicular to the x -axis and the angle is specified with respect to the optical axis z . The size of the reconstructed volume is $23.04 \mu\text{m} \times 23.04 \mu\text{m} \times 23.04 \mu\text{m}$,

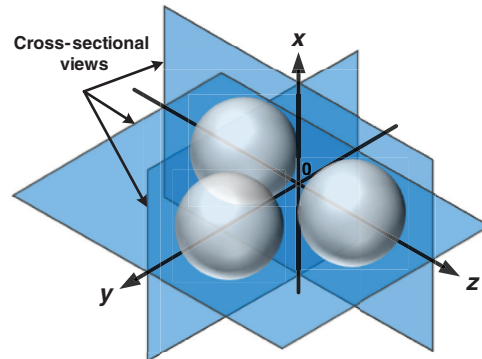


Fig. 3 Simulation geometry comprising three spherical beads with a refractive index difference of 0.03 compared with the background.

with the sampling steps of $\delta x = \delta y = \delta z = 144$ nm. For the network hyperparameters, we choose 600 training iterations in our GPU-based implementation with the batch size of 20, the initial learning rate of 0.001, and the regularization coefficient of $\tau = 1.5$. The reconstructed results by our method and other reconstruction methods are shown in Fig. 4. The SNR defined in Ref. 21 of our result is 25.56 dB, 14 dB higher than the previous works. We can also observe much sharper edges of the reconstructed beads at the interface with less noise in the background from Fig. 5. The comparison between the proposed loss function and other regularized loss functions proves the higher reconstruction quality of the ℓ_1 -norm constraint than the ℓ_2 -case directly.

In addition, we analyze the performance of our method under different noise levels and reduced sampling angles. For the noise test, we add Gaussian noise of different power levels to the 81 simulated measured complex amplitudes, which are represented as different SNRs of the training data. From the curve of the reconstructed SNR versus the noise level, as shown in Fig. 6(a), we can find our method maintains more robustness to the noise than other methods. This is especially useful in the case of shorter exposure time for higher scanning speed, where the data are always readout-noise limited. For the test of reduced sampling angles, we keep the range of incident angles fixed from $-\pi/4$ to $\pi/4$. The total number of the incident angles for the network training decreases from 81. The curve of the reconstructed SNR versus the number of the incident angles is shown in Fig. 6(b). Even with as few as 11 incident angles, we can still achieve comparable performance as the previous methods with 81 angles. This nearly eight-time improvement facilitates the development of high-speed 3-D refractive index imaging.

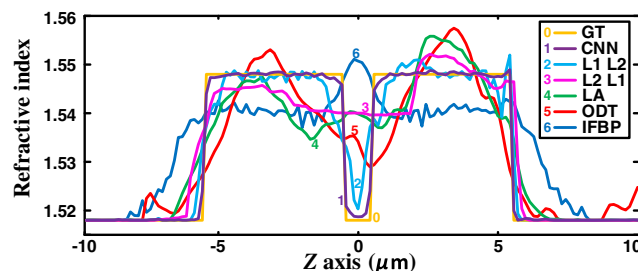


Fig. 5 Comparison of the refractive index profiles along the z -axis reconstructed by different algorithms and the ground truth.

3.2 Tomographic Reconstruction of a Biological Sample

To further validate the capability of the network, we display the experimental results on HeLa cells performed by our tomographic phase microscope as shown in Fig. 1. In detail, we illuminate the HeLa cells in culture medium of refractive index $n_0 = 1.33$ from 41 incident angles evenly distributed from -35° to 35° . The measured hologram with an incident angle of 0° is shown in Fig. 1. The reconstructed volume is $36.86 \mu\text{m} \times 36.86 \mu\text{m} \times 23.04 \mu\text{m}$, composed of $256 \times 256 \times 160$ voxels (with the voxel size of $144 \text{ nm} \times 144 \text{ nm} \times 144 \text{ nm}$). After the selection of hyperparameters, we set the regularization coefficient to $\tau = 5$, with training iterations of 100, the batch size of 20, and the initial learning rate of 0.002. The performance comparison of different methods under different levels of data reduction is shown in Fig. 7. More details can be observed by our method even with fewer incident angles. Moreover, many fewer artifacts and noises exist in our results

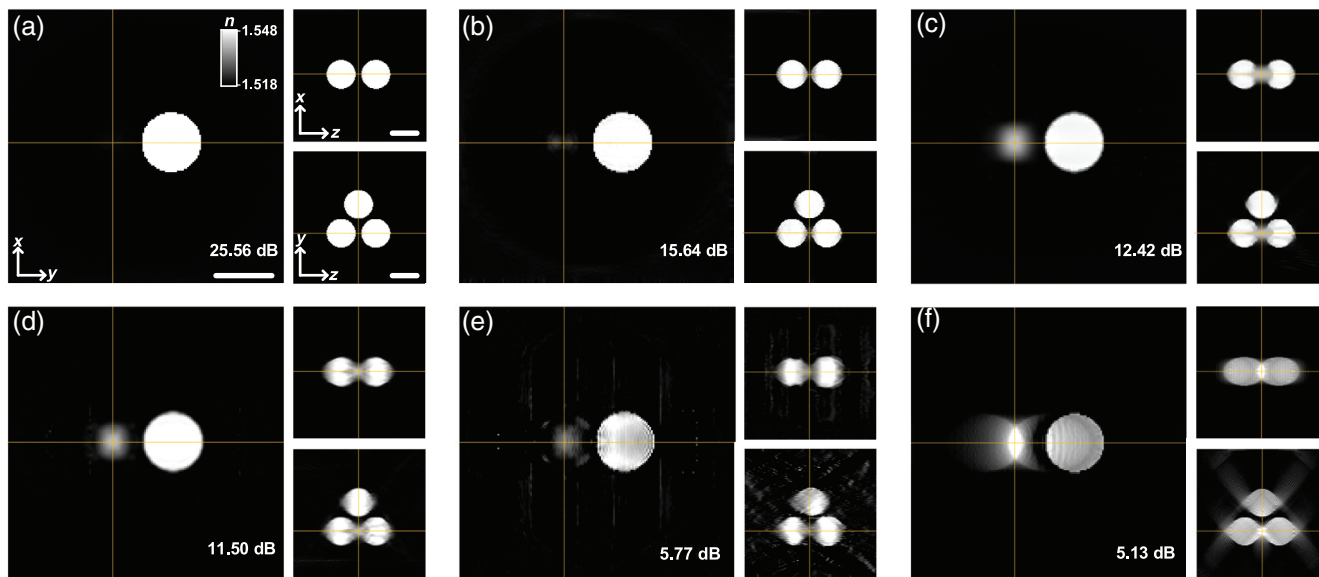


Fig. 4 Reconstruction results of three $5 \mu\text{m}$ beads. Comparison of the cross-sectional slices of the 3-D refractive index distribution of the sample along the $x-y$, $x-z$, and $y-z$ planes reconstructed by (a) proposed CNN, (b) CNN with ℓ_1 fitting, ℓ_2 regularization (L1 L2) and the regularization coefficient of 5, (c) CNN with ℓ_2 fitting, ℓ_1 regularization (L2 L1) and the regularization coefficient of 0.1, (d) learning approach²³ implemented on the same CNN settings (LA) with the regularization coefficient of 0.6, (e) optical diffraction tomography based on the Rytov approximation (ODT)¹³ with the positivity constraint and 100 iterations, and (f) iterative reconstruction based on the filtered backprojection method⁴ with the positivity constraint and 400 iterations. Scale bar, $5 \mu\text{m}$.

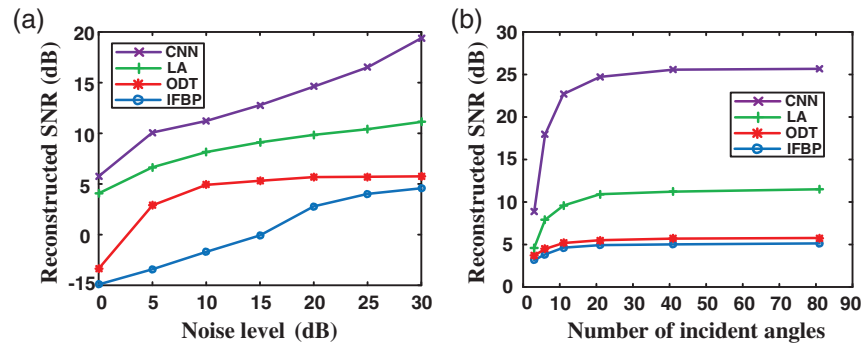


Fig. 6 Performance analysis for proposed approach with the same hyperparameter selection. (a) The curve of the reconstructed SNR versus the noise level and (b) the curve of the reconstructed SNR versus the number of the incident angles.

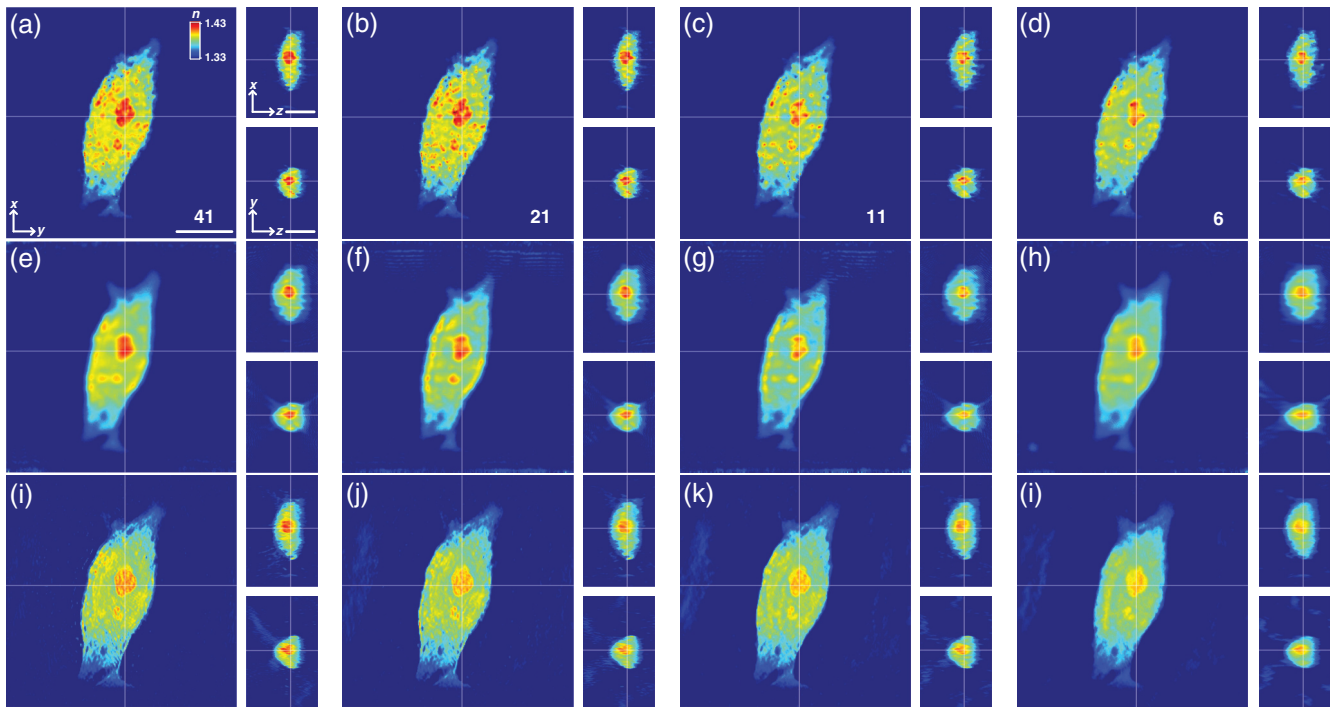


Fig. 7 Comparison of three reconstruction algorithms for various levels of data reduction on a HeLa cell. (a–d) Proposed CNN, (e–h) LA with the regularization coefficient of 1.5, (i–l) ODT with the positivity constraint and 20 iterations, (a, e, and i) 41 training data, (b, f, and j) 21 training data, (c, g, and k) 11 training data, and (d, h, and l) 6 training data. Scale bar, 10 μm .

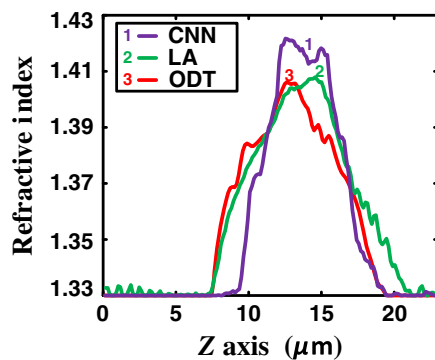


Fig. 8 Comparison of the reconstructed HeLa cell refractive index profiles along the z-axis with 41 training data.

with large data reduction than other methods, which can be seen apparently in Fig. 8.

4 Discussion

4.1 Comparison between L_1 -Norm and L_2 -Norm for Loss Function Design

Loss function design is a crucial element of learning algorithms, which determines the training process of the neural network. Regularized loss function comprises one data-fidelity term and one regularization term.

To the best of our knowledge, the presented study is the first to employ ℓ_1 fitting for the regularized TPM. Generally, the choice of the data-fidelity term depends on the specified noise distribution. However, it is particularly common for solving

normal image restoration problems, as under various constraints, images are always degraded with mixed noise and it is impossible to identify what type of noise is involved. The ℓ_2 -norm fitting relies on strong assumptions on the distribution of noise: there are no heavy tails and the distribution is symmetric. If either of these assumptions fails, then the use of ℓ_2 -norm is not an optimal choice. On the other hand, for the so-called robust formulation based on ℓ_1 -norm fitting, it has been shown that the corresponding statistics can tolerate up to 50% false observations and other inconsistencies.²⁷ Hence, ℓ_1 -norm data-fidelity term relaxes the underlying requirements for the ℓ_2 -case and is well suited to biomedical imaging especially when the noise model is undetermined [as shown in Figs. 4(a)–4(d)] and mixed [as shown in Fig. 6(a)].

As for the regularization term, we finally choose the anisotropic total variation (TV) regularizer in our method, which is an ℓ_1 penalty directly on the image gradient. It is a very strong regularizer, which offers improvements on reconstruction quality to a great extent compared with the isotropic counterpart (ℓ_2 penalty).²⁹ Therefore, the edges are better preserved, which can be seen apparently from the comparison between Figs. 4(a) and 4(b).

4.2 Selection of Hyperparameters

Selection of hyperparameters has an important effect on tomographic reconstruction results. In practice, many learning algorithms involve hyperparameters (10 or more), such as initial learning rate, minibatch size, and regularization coefficient. Reference 31 introduces a large number of recommendations for training feed-forward neural networks and choosing the multiple hyperparameters, which can make a substantial difference (in terms of speed, ease of implementation, and accuracy) when it comes to putting algorithms to work on real problems. Unfortunately, optimal selection of hyperparameters is challenging due to the high computational cost when using traditional regularized iterative algorithms.^{21,23}

In this study, our GPU-based implementation of CNN runs computation-intensive simulations at low cost and is possible to make the optimization of hyperparameters a more reproducible and automated process with modern computing facilities. Thus, we can gain better and more robust reconstruction performance with the GPU-based learning method. During the simulation and experiment, selection of hyperparameters varies with the biological sample and the range of incident angles. To make a convincing comparison, optimal hyperparameters have been selected for all the other reconstruction methods. The refractive index difference $\delta n(\mathbf{r})$ is initialized with a constant value of 0 for all the methods, and different optimal regularization coefficients are chosen for different regularized loss functions due to the different combinations of data-fidelity term and regularization term. The number of iterations is set to guarantee the convergence of each method, as shown in Fig. 9.

4.3 Subgradient Method and Adam Algorithm

In convex analysis,³⁰ the subgradient generalizes the derivative to functions that are not differentiable. A vector $g \in \mathbb{R}^n$ is a subgradient of a convex function at x if

$$f(y) \geq f(x) + g^T(y - x) \quad \forall y. \quad (5)$$

If f is convex and differentiable, then its gradient at x is a subgradient. But a subgradient can exist even when f is not

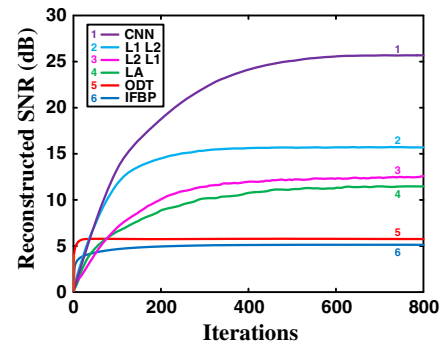


Fig. 9 Reconstructed SNR plotted as a function of the number of iterations for different reconstruction methods on simulated data. Hyperparameters are declared in Sec. 3.1.

differentiable at x . There can be more than one subgradient of a function f at a point x . The set of all subgradients at x is called the subdifferential, and is denoted by $\partial f(x)$. Considering the absolute value function $|x|$, the subdifferential is

$$\partial|x| = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \\ [-1, 1], & x = 0 \end{cases}. \quad (6)$$

Subgradient methods are subgradient-based iterative methods for solving nondifferentiable convex minimization problems.

Adam is an algorithm for first-order (sub)gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The method is aimed toward machine learning problems with large datasets and/or high-dimensional parameter spaces. The method is also appropriate for nonstationary objectives and problems with very noisy and/or sparse gradients. Adam works well in practice and compares favorably with other stochastic optimization methods regarding the computational performance and convergence rate.²⁶ It is straightforward to implement, is computationally efficient, and has little memory requirements, which is robust and well suited to TPM. Compared with the stochastic proximal gradient descent (SPGD) algorithm reported in Ref. 21, our GPU-based CNN trained with the Adam algorithm for optimization converges to the same SNR level and achieves twice the rate of convergence as shown in Fig. 10. To show the higher convergence rate of Adam fairly, we use the proposed ℓ_1 -norm loss function

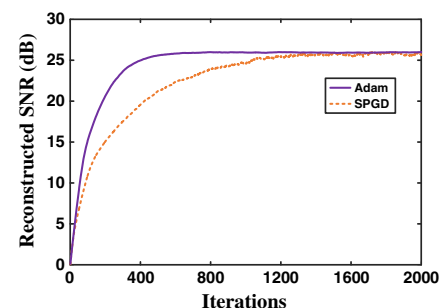


Fig. 10 Reconstructed SNR of proposed approach plotted as a function of the number of iterations for two different training optimization algorithms on simulated data with the same hyperparameters declared in Sec. 3.1.

for both the Adam and SPGD training processes here, thus producing the same reconstructed SNR after convergence.

5 Conclusion

We have demonstrated a GPU-based implementation of deep CNN to model the propagation of light in inhomogeneous sample for TPM and have applied it to both synthetic and biological samples. The experimental results verify its superior reconstruction performance over other tomographic reconstruction methods, especially when we take fewer measurements. Furthermore, our CNN is much more general under different optical systems and arbitrary illumination patterns as its design is illumination-independent. Importantly, this approach can not only enlarge the applications of optical tomography in biomedical imaging, but also open rich perspectives for the potential of deep neural networks in the optical society.

Disclosures

The authors have no relevant financial interests in the article and no other potential conflicts of interest to disclose.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61327902 and 61671265). The authors gratefully acknowledge Ulugbek S. Kamilov, Alexandre Goy, and Demetri Psaltis for providing the code and their helpful suggestions.

References

1. F. Zemike, "How I discovered phase contrast," *Science* **121**(3141), 345–349 (1955).
2. G. Popescu, *Quantitative Phase Imaging of Cells and Tissues*, McGraw Hill Professional, New York (2011).
3. E. Cuche, F. Bevilacqua, and C. Depeursinge, "Digital holography for quantitative phase-contrast imaging," *Opt. Lett.* **24**(5), 291–293 (1999).
4. W. Choi et al., "Tomographic phase microscopy," *Nat. Methods* **4**(9), 717–719 (2007).
5. Y. Cotte et al., "Marker-free phase nanoscopy," *Nat. Photonics* **7**(2), 113–117 (2013).
6. T. Kim et al., "White-light diffraction tomography of unlabelled live cells," *Nat. Photonics* **8**(3), 256–263 (2014).
7. S. Y. Lee et al., "The effects of ethanol on the morphological and biochemical properties of individual human red blood cells," *PLoS One* **10**(12), e0145327 (2015).
8. J. Yoon et al., "Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning," *Sci. Rep.* **7**, 6654 (2017).
9. S. Shin et al., "Active illumination using a digital micromirror device for quantitative phase imaging," *Opt. Lett.* **40**(22), 5407–5410 (2015).
10. D. Jin et al., "Tomographic phase microscopy: principles and applications in bioimaging [invited]," *J. Opt. Soc. Am. B* **34**(5), B64–B77 (2017).
11. D. Jin et al., "Dynamic spatial filtering using a digital micromirror device for high-speed optical diffraction tomography," *Opt. Express* **26**(1), 428–437 (2018).
12. J. Lim et al., "Comparative study of iterative reconstruction algorithms for missing cone problems in optical diffraction tomography," *Opt. Express* **23**(13), 16933–16948 (2015).
13. Y. Sung et al., "Optical diffraction tomography for high resolution live cell imaging," *Opt. Express* **17**(1), 266–277 (2009).
14. M. M. Bronstein et al., "Reconstruction in diffraction ultrasound tomography using nonuniform fft," *IEEE Trans. Med. Imaging* **21**(11), 1395–1401 (2002).
15. Y. Sung and R. R. Dasari, "Deterministic regularization of three-dimensional optical diffraction tomography," *J. Opt. Soc. Am. A* **28**(8), 1554–1561 (2011).
16. E. Wolf, "Three-dimensional structure determination of semi-transparent objects from holographic data," *Opt. Commun.* **1**(4), 153–156 (1969).
17. A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, Society for Industrial and Applied Mathematics, Philadelphia (2001).
18. K. Lee et al., "Time-multiplexed structured illumination using a DMD for optical diffraction tomography," *Opt. Lett.* **42**(5), 999–1002 (2017).
19. V. Katkovnik et al., "Computational super-resolution phase retrieval from multiple phase-coded diffraction patterns: simulation study and experiments," *Optica* **4**(7), 786–794 (2017).
20. J. W. Goodman, *Introduction to Fourier Optics*, Roberts and Company Publishers, Greenwood Village (2005).
21. U. S. Kamilov et al., "Optical tomographic image reconstruction based on beam propagation and sparse regularization," *IEEE Trans. Comput. Imaging* **2**(1), 59–70 (2016).
22. U. S. Kamilov et al., "A recursive born approach to nonlinear inverse scattering," *IEEE Signal Process. Lett.* **23**(8), 1052–1056 (2016).
23. U. S. Kamilov et al., "Learning approach to optical tomography," *Optica* **2**(6), 517–522 (2015).
24. H. B. Demuth et al., *Neural Network Design*, Martin Hagan (2014).
25. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
26. D. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).
27. T. Kärkkäinen, K. Kunisch, and K. Majava, "Denoising of smooth images using l1-fitting," *Computing* **74**(4), 353–376 (2005).
28. E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory* **52**(2), 489–509 (2006).
29. X. Jin et al., "Anisotropic total variation for limited-angle CT reconstruction," in *IEEE Nuclear Science Symp. and Medical Imaging Conf.*, pp. 2232–2238, IEEE (2010).
30. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey (2015).
31. G. Montavon, G. B. Orr, and K.-R. Müller, *Neural Networks: Tricks of the Trade*, Springer, Berlin, Heidelberg (2012).

Hui Qiao is a PhD candidate at the Department of Automation, Tsinghua University, Beijing, China. His research interests include biomedical imaging and deep learning.

Qionghai Dai is currently a professor at the Department of Automation, Tsinghua University, Beijing, China. He received his PhD from Northeastern University, Liaoning province, China, in 1996. His research interests include microscopy imaging for life science, computational photography, computer vision, and 3-D video.

Biographies for the other authors are not available.