

Single-frame label-free cell tomography at speed of more than 10,000 volumes per second: supplementary material

Baoliang Ge^{1,2,†}, Yanping He^{3,†}, Mo Deng^{1,†}, Md Habibur Rahman³, Yijin Wang³, Ziling Wu¹, Chung Hong N. Wong^{4,5}, Michael K. Chan^{4,5}, Yi-Ping Ho^{3,5,6}, Liting Duan³, Zahid Yaqoob², Peter T. C. So^{1,2,7,8,*}, George Barbastathis^{1,8,9,*}, and Renjie Zhou^{3,*}

¹Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Laser Biomedical Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³Department of Biomedical Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

⁴School of Life Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

⁵Centre for Novel Biomaterials, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

⁶Hong Kong Branch of CAS Center for Excellence in Animal Evolution and Genetics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China

⁷Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁸Interdisciplinary Research Group on Critical Analytics for Manufacturing Personalized medicine (CAMP), Singapore-MIT Alliance for Research and Technology Centre, Singapore

⁹Intra-Create Thematic Grant on Retinal Analytics through Machine learning aiding Physics (RAMP), Singapore-MIT Alliance for Research and Technology Centre, Singapore

[†]These authors contributed equally to this work

1. Mathematical formulation for Phase Approximant retrieval

The irradiance of the interference between the sample field, resulting from illumination with N illumination angles, and the reference field can be described as:

$$I(\mathbf{r}) = \left| \sum_{m=1}^N E_S^{(m)}(\mathbf{r}) + E_R \right|^2 = |E_R|^2 + \sum_{m=1}^N |E_S^{(m)}(\mathbf{r})|^2 + \sum_{m=1}^N \langle E_S^{(m)}(\mathbf{r}), E_R \rangle + \sum_{i,j(i \neq j)} \langle E_S^{(i)}(\mathbf{r}), E_S^{(j)}(\mathbf{r}) \rangle, \quad (S1)$$

where $E_S^{(m)}(\mathbf{r})$ is the sample complex field due to the m^{th} illumination angle, E_R is the reference field that is a plane wave, and $\langle \cdot \rangle$ denotes the spatial correlation function that describes the interference between fields. The 3rd and 4th terms in Eqn. S1 denote the interference between the scattered beams and the reference beam, which will show up in the +1st and -1st orders in the Fourier domain. Both the +1st and the -1st orders contain the same sample structural information. Next, we expand the +1st order (i.e., the 3rd term in Eqn. S1) as:

$$\sum_{m=1}^N \langle E_S^{(m)}(\mathbf{r}), E_R \rangle = \sum_{m=1}^N |E_S^{(m)}(\mathbf{r})| |E_R| \cos \left(\left(\mathbf{k}_S^{(m)} + \mathbf{k}_R \right) \cdot \mathbf{r} + \Delta\phi^{(m)}(\mathbf{r}) \right), \quad (S2)$$

where $\mathbf{k}_S^{(m)}$ is the wave-vector of the scattered field, corresponding to the m^{th} illumination angle, and \mathbf{k}_R is the wave-vector of the reference beam. Applying a 2D spatial Fourier transform over the $+1^{\text{st}}$ order as described in Eqn. S2, we have:

$$\tilde{I}_{+1}(\mathbf{k}) = |E_R| \sum_{m=1}^N F \left[\left| E_S^{(m)}(\mathbf{r}) \right| \exp \left(i \Delta \phi^{(m)}(\mathbf{r}) \right) \right] * \delta(\mathbf{k} - \mathbf{k}_S^{(m)} - \mathbf{k}_R), \quad (\text{S3})$$

where $\tilde{I}_{+1}(\mathbf{k})$ denotes the $+1^{\text{st}}$ order spectrum, $F[\cdot]$ denotes the Fourier transform operator, and $*$ denotes the 2D convolution in space. For the m^{th} illumination angle, the original sample spectrum is shifted by $\mathbf{k}_S^{(m)} + \mathbf{k}_R$. Therefore, four bright regions in the $+1^{\text{st}}$ order are observed as shown in Fig. 5b. In experiments, the numerical aperture (NA) of the objective lens 2 (OL2) limits the bandwidth of the spectrum that can be used. Therefore, a pupil function is added to the $+1^{\text{st}}$ order spectrum $\tilde{I}_{+1}(\mathbf{k})$ as:

$$\tilde{I}'_{+1}(\mathbf{k}) = P(\mathbf{k} - \mathbf{k}_R) \tilde{I}_{+1}(\mathbf{k}), \quad (\text{S4})$$

where $P(\mathbf{k} - \mathbf{k}_R)$ is the shifted pupil function with $P(\mathbf{k})$ defined as follows: $P(\mathbf{k}) = 1$ when $|\mathbf{k} - \mathbf{k}_R| \leq k_c$; and $P(\mathbf{k}) = 0$, when $|\mathbf{k} - \mathbf{k}_R| > k_c$. $k_c = \frac{2\pi}{\lambda} NA$ is the cut-off spatial frequency of the pupil function, λ is the wavelength, and $NA = 1.25$ is the numerical aperture of OL2. In the Fourier domain, we have N sample spectra, corresponding to N illumination angles, which overlap and cannot be easily separated through linear filtering. In our experiments with $N=4$, we noticed that the sample spectrum of each illumination angle is mainly concentrated in the low-frequency band. Therefore, there is a possibility to separate each sample spectrum without suffering severe aliasing. For the m^{th} illumination angle, following a spectral filtering process, its corresponding estimated sample field $\hat{E}_S^{(m)}(\mathbf{r})$ can be described as:

$$\begin{aligned} \hat{E}_S^{(m)}(\mathbf{r}) &= \left| \hat{E}_S^{(m)}(\mathbf{r}) \right| \exp \left(i \left(\Delta \hat{\phi}^{(m)}(\mathbf{r}) + (\mathbf{k}_R + \mathbf{k}_S^{(m)}) \cdot \mathbf{r} \right) \right) \\ &= F^{-1} \left[p^{(m)}(\mathbf{k} - \mathbf{k}_R - \mathbf{k}_S^{(m)}) \tilde{I}'_{+1}(\mathbf{k}) \right], \end{aligned} \quad (\text{S5})$$

where $F^{-1}[\cdot]$ denotes the inverse Fourier transform, and $p^{(m)}(\mathbf{k})$ is the spectral filter that selects the information for the m^{th} illumination angle. We define $p^{(m)}(\mathbf{k}) = 1$, when $|\mathbf{k} - \mathbf{k}_R - \mathbf{k}_S^{(m)}| < k_f$, where k_f is the cut-off spatial frequency; and $p^{(m)}(\mathbf{k}) = 0$, when $|\mathbf{k} - \mathbf{k}_R - \mathbf{k}_S^{(m)}| \geq k_f$. The determination of k_f value presents a trade-off: a larger k_f value can retain the high spatial frequencies from the sample at the cost of introducing more aliasing, whereas a smaller k_f can reduce aliasing by sacrificing high spatial frequencies from the sample. With an additional measurement of the background (the sample-free field of view, whose Fourier transform at the $+1^{\text{st}}$ order is represented as $\tilde{I}_{BG}(\mathbf{k})$), the phase map corresponding to the m^{th} illumination angle is retrieved as:

$$\Delta \hat{\phi}^{(m)}(\mathbf{r}) = \angle \left[F^{-1} \left[p^{(m)}(\mathbf{k} - \mathbf{k}_R - \mathbf{k}_S^{(m)}) \tilde{I}_m(\mathbf{k}) \right] \right]$$

$$-\angle\left[F^{-1}\left[p^{(m)}(\mathbf{k}-\mathbf{k}_R-\mathbf{k}_S^{(m)})\tilde{I}_{BG}(\mathbf{k})\right]\right], \quad (\text{S6})$$

where \angle denotes the phase of a complex field. The phase maps retrieved from the interferogram with multiplex illumination are termed ‘Phase Approximants’. In our angle-multiplex interferometric microscope (Fig. 5a), we select four illumination angles that are uniformly distributed along the outer circle of the back focal plane (BFP) of objective lens 1 (OL1) with the same elevation angle of around 59.27° as illustrated by the red spots in Fig. S1a. We found that these four angles are sufficient for reconstructing a high-quality 3D RI map when using the DNN-based reconstruction model (refer to Section 2.2 in the main text). To balance between preserving high-frequency components and minimizing spatial frequency aliasing, we need to optimize the spectral filter size k_f . Here we use k_c as the unit to characterize the size of the spectral filter Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC) to quantify the accuracy of the Phase Approximants. PCC and MAE are defined as:

$$\text{PCC}(\phi, \hat{\phi}) = \frac{\sum_i (\phi_i - \bar{\phi})(\hat{\phi}_i - \bar{\hat{\phi}})}{\sqrt{\sum_i (\phi_i - \bar{\phi})^2 \sum_i (\hat{\phi}_i - \bar{\hat{\phi}})^2}}, \quad (\text{S7})$$

and

$$\text{MAE}(\phi, \hat{\phi}) = \frac{1}{N} \sum_i |\phi_i - \hat{\phi}_i|, \quad (\text{S8})$$

where ϕ is the true phase map retrieved with the angle-scanning scheme; $\hat{\phi}$ is the Phase Approximant extracted from the single-frame multiplex interferogram; i denotes the index of each pixel and N is the total number of pixels. In Fig. S1 b-c, we show the Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC) values when k_f is $0.3 - 0.8k_c$. Note that a small spectral filter size of $0.3 k_c$ or less incurs a large loss of high-frequency information that causes a great loss of phase retrieval accuracy. On the other hand, it is observed that a large filter size of $0.8k_c$ deteriorates the accuracy of the Phase Approximants as expected due to the aliasing problem, even though less high-frequency information is then cut off. Therefore, as a reasonable compromise we choose the spectral filter size to be $0.5 k_c$, which is expected to well balance between preserving high-frequencies and avoiding aliasing.

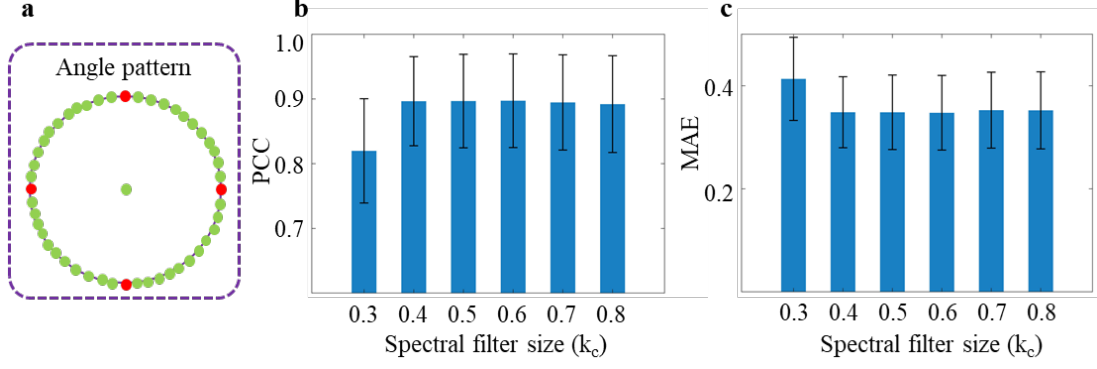


Figure S1. Determination of the optimum spectral filter size for retrieving the Phase Approximants of each illumination angle. (a) Illustration of the illumination angle pattern at the back focal plane of objective lens 1 (OL1). The green and red dots show all the illumination angles used in the angle-scanning scheme, while the red dots show the four illumination angles used in the angle-multiplexing scheme. (b) The PCC and (c) MAE values when the spectral filter size k_f is 0.3 – 0.8 k_c , respectively.

2. 3D RI map reconstruction models

Here we introduce the physical principles and mathematical framework for reconstructing 3D RI maps with the angle scanning interferometric microscope. It is worth noting that when using data-driven methods to improve the imaging performance in SILACT, a reliable ground truth 3D RI map reconstruction model, based on accurate physical modelling, is required to construct the training dataset. Firstly, we formulate the forward model for reconstructing the 3D RI map from the measured complex fields corresponding to different illumination as follows:

$$g_m = S_m(\mathbf{x}), \quad (\text{S9})$$

where g_m is the measured complex field of the m^{th} illumination angle, \mathbf{x} is the 3D RI map that needs to be solved, and S is the operator that maps the 3D RI map to the measured complex field. S could be formulated with different physical models. After that, the inverse problem of Eqn. S9 can be described as following:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \sum_{m=1}^N \|S_m(\mathbf{x}) - g_m\|_2^2, \quad (\text{S10})$$

where $\hat{\mathbf{x}}$ is the estimated 3D RI map when we have N illumination angles. As the forward model S is nonlinear, directly solving Eqn. S10 is very difficult. One of the solutions is to linearize Eqn. S10 with a proper approximation. For example, when the scattering of the light in the specimen is weak, we can simplify Eqn. S9 as follows(I , 2):

$$S_m(\mathbf{x}) = A_m \mathbf{x} = \left(\frac{k^2}{4\pi} \right) \int_V \left(\frac{\mathbf{x}^2(\mathbf{r}')}{n_m^2} - 1 \right) e^{-i(k-k_m)\mathbf{r}'} d\mathbf{r}', \quad (\text{S11})$$

where A_m represents the forward model that maps the 3D RI map into the 3D Fourier space for the m^{th} illumination angle, \mathbf{r}' denotes the spatial coordinate of a point in the observed object, $\mathbf{k} = (k_x, k_y)$ is the Fourier domain variable, and $\mathbf{k}_m = k\mathbf{s}_m = k(s_x^{(m)}, s_y^{(m)}, s_z^{(m)})$ is the wave-vector of the m^{th} illumination field. When imaging living cells, one can apply the first-order Rytov approximation(1), so that the left-hand side of Eqn. S9 can be rewritten as:

$$g_m = \frac{ks_z^{(m)}}{2\pi i} \int \int_{-\infty}^{\infty} U_0^{(m)}(\mathbf{r}) \ln \frac{U(\mathbf{r})}{U_0^{(m)}(\mathbf{r})} e^{-i(k s_x^{(m)} x + k s_y^{(m)} y)} dx dy, \quad (\text{S12})$$

where $U_0^{(m)}(\mathbf{r})$ is the m^{th} illumination field, $U(\mathbf{r}) = U_0^{(m)}(\mathbf{r}) + U_s(\mathbf{r})$, and $U_s(\mathbf{r})$ is the sample scattered field. Therefore, a linear forward model is established. However, the inverse problem is still difficult to solve since it is ill-posed due to the missing-cone problem (missing axial frequency supports at low lateral frequencies of the 3D Fourier space) and thus likely to miss structural details and amplify noise artifacts(2). To mitigate those problems, a regularization term can be added to Eqn. S10 (this method is also called regularized ODT(2)), then the inverse problem becomes:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmin}} \sum_{m=1}^N \|A_m \mathbf{x} - g_m\|_2^2 + \alpha R(\mathbf{x}), \quad (\text{S13})$$

where α is the regularization coefficient, and $R(\mathbf{x})$ is the regulariser (also referred to as ‘the prior’). The aforementioned forward and inverse models are the basic principle of ODT with Rytov approximation. To simplify our description, in this paper, we called this method Rytov ODT; the direct inverse of Eqn. S9 is what we called Rytov ODT without regularization (ROWOR); and we called the method of using Eqn. S13 to solve the 3D RI maps as Rytov ODT with regularization (ROWR).

Note that the presented linear model may not be accurate as it assumes single scattering in the biological specimen. For more accurate 3D RI reconstructions, a nonlinear forward model that models multiple scattering is needed. U. S. Kamilov *et al* proposed a nonlinear forward model based on the beam propagation method (BPM), which demonstrated a more accurate reconstruction of the 3D RI maps of cell (3, 4). In the BPM model, the 3D object is sliced along the z -direction into $k = 1, 2, \dots, K$ layers, which assemble the structures of neural networks. The optical field between layers is propagated via a linear mapping:

$$S_k(\mathbf{x}) = \text{diag}(p_k(\mathbf{x}_k)) H S_{k-1}(\mathbf{x}), \quad (\text{S14})$$

where $S_k(\mathbf{x})$ is the optical field of the k^{th} layer, and $\text{diag}(\cdot)$ is the operator that converts a vector into a diagonal matrix. H is a diffraction operator which is simplified for scalar paraxial (also known as Fresnel) propagation, $p_k(\mathbf{x}_k) = \exp(ik\mathbf{x}_k\delta z)$ describes the refraction taking place in the k -th layer, and δz denotes the spacing between layers. Then, the inverse problem can be formulated as:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \sum_{m=1}^N \left\| S_K^{(m)}(\mathbf{x}) - g_m \right\|_2^2, \quad (\text{S15})$$

where g_m is the measured optical field in the spatial domain, which is different from what we used in Eqn. S13. Here $g_m(\mathbf{r}') = E_m(\mathbf{r}') \exp(i\phi_m(\mathbf{r}'))$, where $E_m(\mathbf{r}')$ is the spatial distribution of the amplitude of the measured optical field, and $\phi_m(\mathbf{r}')$ is the measured phase map. The strategy to solve this problem is based on gradient descent. Firstly, we calculate the gradient of the loss function by iterating between different layers, then we update the RI values of different layers by applying total variation (TV) regularization. The optimization process in this method is similar to training a neural network, therefore, it is also called Learning Tomography based on Beam Propagation Method (LT-BPM) (4).

3. Assessment of the LT-BPM reconstructed RI maps as the ground truth

As a 3D RI reconstruction algorithm (3, 4), LT-BPM can retrieve the 3D RI maps with good accuracy. Here we reconstruct the 3D RI map of a polystyrene bead (RI ~ 1.595 , diameter $\sim 10 \mu\text{m}$) with the ROWOR, ROWR, and LT-BPM methods, respectively. The beads were immersed in an index-matching liquid solution with RI=1.56 and sandwiched with two No. 1.5 coverslips. The 3D RI maps are reconstructed based on the interferograms acquired from 49 illumination angles using our interferometric microscope apparatus. Note that to accelerate the convergence, we initiated the LT-BPM method with the reconstruction results from the ROWOR method. The comparison of reconstruction results is shown in Fig. S2a-c, respectively. In Fig. S2 f-g, we quantify the difference between the ground truth (i.e., assuming a perfect spherical bead) and the reconstructed results by devising a metric based on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The formula for calculating RMSE is defined as:

$$\text{RMSE}(n, \hat{n}) = \frac{1}{N} \sqrt{\sum_i |n_i - \hat{n}_i|^2}, \quad (\text{S16})$$

where n and \hat{n} are the ground truth and reconstructed 3D refractive index maps, respectively. Noticeably, in ROWR and LT-BPM the value of the regularization coefficient α can be adjusted to balance between the data-fidelity term and the regularization term to denoise and to mitigate the missing-cone issues. Usually, this parameter tuning process is empirical. To fairly compare the 3D RI maps, the regularization coefficient α is chosen to be 5×10^{-5} , and each algorithm (Rytov with regularization and LT-BPM) iterates 50 times.

In Fig. S2 d-e, we show the lateral and axial cross-sections of the bead's RI profiles along the directions indicated by the white dashed lines in Fig.S2 a-c. It is found that, despite ROWOR is fast in reconstruction, it incurs two major artifacts: first, some of the region in the reconstruction has RI values lower than 1.56 that deviates from the real situations; second, the RI reconstructions exhibit an elongation along the axial

direction that is attributable to the missing-cone problem as mentioned earlier, which causes the reconstructed RI values to be under-estimated (1, 2, 5). We found that the non-negativity constraint in ROWR or LT-BPM is sufficient to resolve the first type of artifacts. However, LT-BPM is better than the ROWR at mitigating the axial elongation artifact, as indicated in Fig. S2e. Since ROWR and LT-BPM share the same regularization coefficient, we conclude that the improved performance is resulted from the introduction of the nonlinear scattering model (i.e., the BPM model). To further confirm the conclusions, we imaged 100 polystyrene beads and calculated the MAE and RMSE between the ground truth and the RI maps reconstructed by the same reconstruction methods. From the results shown in Fig. S2f-g, we conclude that LT-BPM indeed outperforms the other two reconstruction methods.

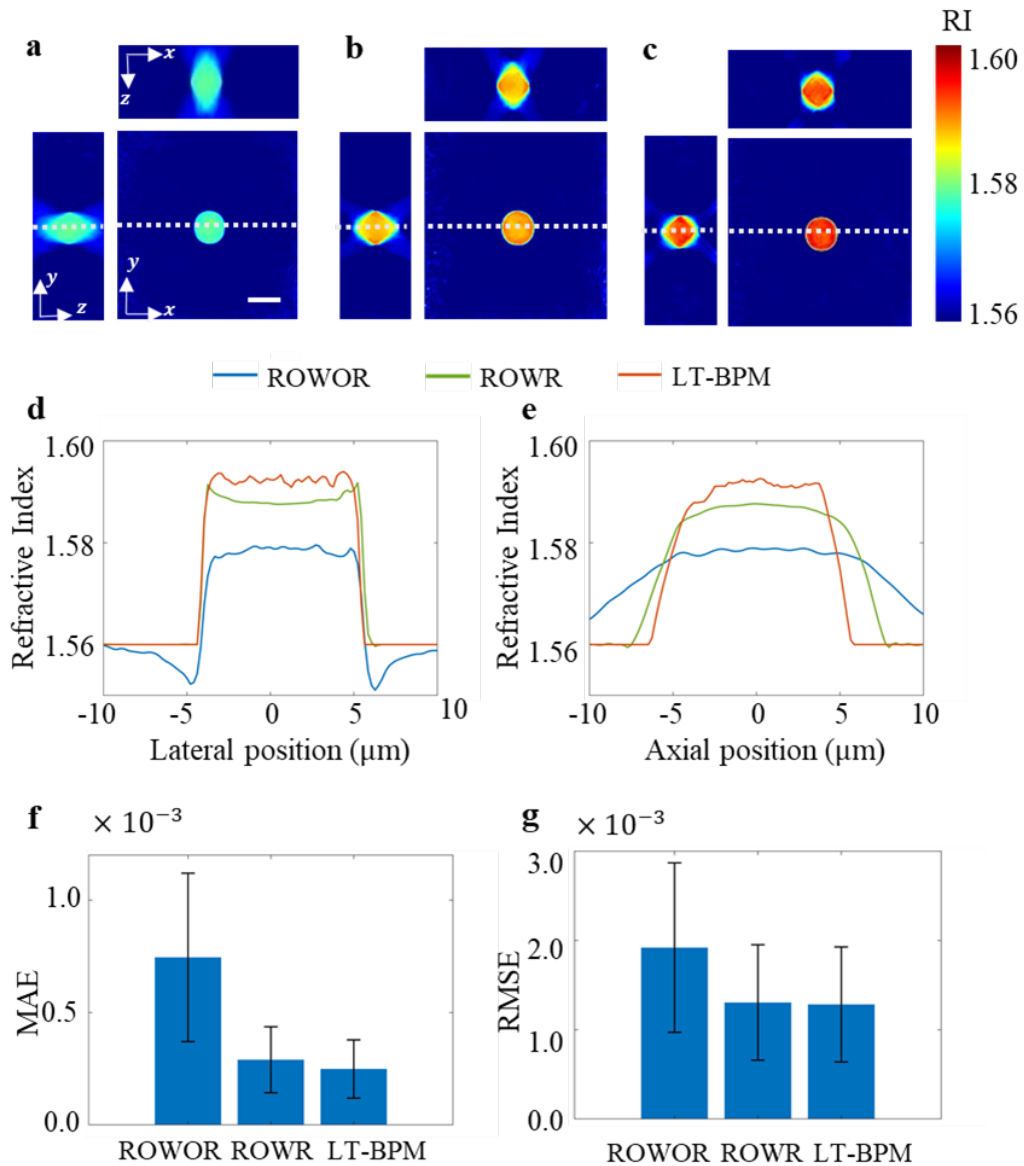


Figure S2. Comparison of different 3D RI reconstruction methods on a standard bead. (a)-(c), cross-sections of the 3D RI maps of the bead reconstructed with ROWOR, ROWR, and LT-BPM, respectively. (d) and (e), the RI profiles from different reconstruction methods along the lateral and axial position indicated by the white dashed lines. (f) and (g), quantitative evaluations of different reconstruction

models' performance (i.e., ROWOR, ROWR, and TT-BPM) using MAE and RMSE. ROWOR stands for Rytov ODT without regularization, while ROWR is the short format of Rytov ODT with regularization. The scale bar in (a) denotes $10\mu\text{m}$.

Even if the accuracy of the ground truth RI maps reconstructed with LT-BPM has been validated with polystyrene beads, one may still need further validations on cells that usually have complex structures. As we do not know the structures of real cells, here we used an artificial cell phantom sample fabricated with 3D printing through two-photon polymerization (6) to evaluate the performance of ROWOR, ROWR, and LT-BPM. The 3D RI maps reconstructed with the aforementioned methods are shown in Fig. S3a-c, while the ground truth RI map is shown in Fig. S3d (note that the cell phantom is immersed in an index-matching liquid solution with RI of 1.50). The left column shows the y - z cross-sections of the 3D RI maps, while the right three columns show the x - y cross-sections at different layers at $z = 0.2\mu\text{m}$, $1\mu\text{m}$, and $1.8\mu\text{m}$. Compared with the ground truth, we can see that LT-BPM is capable of recovering more details and retrieving correct RI values than the other two reconstruction methods. The results are also validated in Fig. S3e, where the MAE, RMSE, and PCC values are calculated between the reconstructed RI maps and the ground truth. From the evaluation results, we can conclude that LT-BPM method has performed ROWOR and ROWR methods.

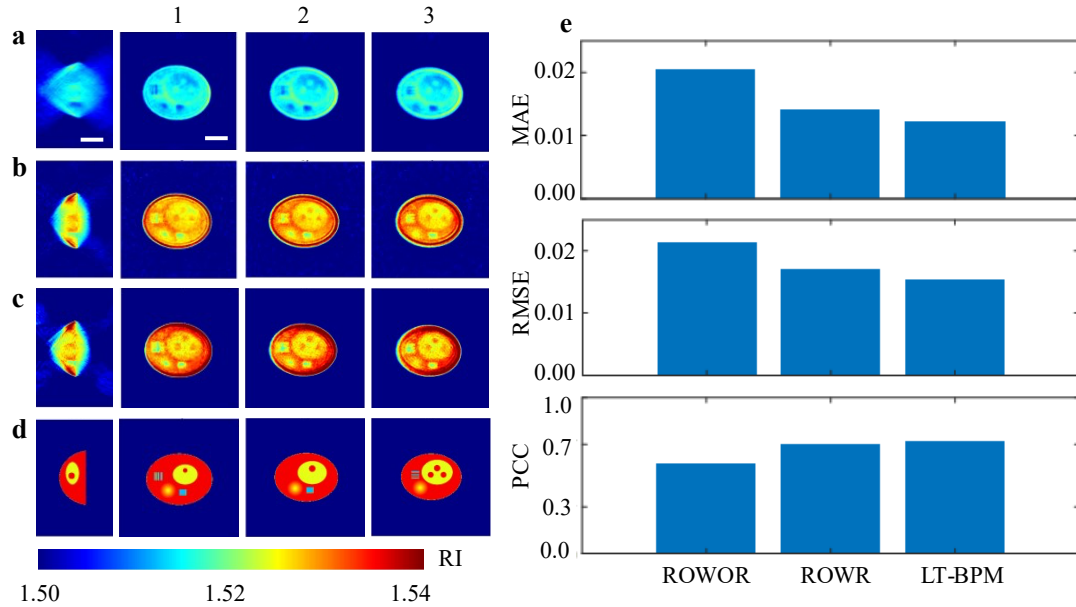


Figure S3. Comparison of different 3D RI reconstruction methods on a cell phantom. (a)-(c), the reconstructed 3D RI maps of the cell phantom with ROWOR, ROWR and LT-BPM, respectively. The cross-section of the y - z plane and the x - y cross-sections at the z locations (1) $0.2\mu\text{m}$, (2) $1\mu\text{m}$, (3) $1.8\mu\text{m}$ (the bottom layer of the cell phantom is set to $z=0\mu\text{m}$) are shown. (d) The ground truth 3D RI maps of the cell phantom. (e) Quantitative evaluations of different reconstruction methods' performance with MAE, RMSE, and PCC when compared with the ground RI maps. The scale bars in (a) denote $10\mu\text{m}$.

4. RI reconstruction schemes

In Section 2.2 of the main text, we compared the reconstructions from three different schemes, i.e., LT-BPM, MDODT, and SILACT. Here, we introduce each scheme in more depth. The number of image acquisitions is denoted as K in this section.

- (i) LT-BPM reconstructions based on K image acquisitions. In this scheme, interferograms based on K illumination angles are individually captured. The LT-BPM algorithm takes in these K phase maps, each extracted from the corresponding single-angle interferogram, and produces the final reconstructions iteratively. Since the phase estimation from the single angle interferograms is highly reliable, we will refer to phase maps thus estimated as true phase maps.
- (ii) Multi-frame deep-learning ODT (MDODT) with K illumination angles. The data acquisition process is identical to that in scheme (i), but a machine learning engine is trained to map K true phase maps to 3D refractive index (RI) maps. It costs the same acquisition time as that of LT-BPM for the same K . Though the training may take a few hours, the inference time is negligible once the network has been trained. Note that here, for $K = 1$, it refers to the scenario where only the normal incident illumination is used.
- (iii) Single-frame deep-learning ODT (a.k.a. the reconstruction model in SILACT), where the input to the pipeline is the single-shot angle-multiplex interferogram based on 4 uniformly spaced illumination angles (but only 1 acquisition). The machine learning engine is trained to produce 3D RI maps from the Phase Approximants of the four individual phase maps.

5. Architecture of U-Net with residual blocks and training specifics

As mentioned in *Methods*, the U-Net with residual blocks is used as the fundamental architecture for DNN-L, DNN-H, and DNN-S, illustrated earlier in Fig. 5c. The architecture of the U-Net architecture with residual blocks is shown in Fig. S4a, where we design a U-Net structure consisting of multiple down-residual blocks (DRB), gradually extracting features from the input, followed by several up-residual blocks (URB) to scale the feature maps up to the correct size. Skip connections are proven beneficial for the spatial resolution of the reconstructions. In the meanwhile, we use residual blocks in the U-Net to stabilize training and avoid gradient exploding/vanishing when the network depth is large (7). The detailed structures of DRB, RB, and URB in our U-Net with residual blocks architecture are shown in Fig. S4b.

The major ML engine used for the validation and demonstration of the 3D imaging cytometer is trained with a dataset that only contains ~ 900 input-ground truth pairs of NIH/3T3 cells. Since the RBCs have structures distinct from eukaryotic cells, in the section for observing the deformation of RBCs we also tested the RBCs in the microfluidic channels with a ML engine trained on ~ 500 input-ground truth pairs of RBCs. In Section 2 of the main text, the sizes of the testing datasets for NIH/3T3 cell, HEK293T cells, HeLa cells, and COS-7 cells are 39, 69, 85, and 68, separately. The

training is conducted on a Nvidia Tesla K80 GPU using the open-source machine learning Platform TensorFlow & Pytorch. In each training of DNN-L, DNN-H, and DNN-S, we trained 500 epochs with a batch size of 2 and a learning rate of 0.01. We used Adaptive Moment Estimation (Adam) optimizer, and in each epoch, 5% of the data is used for validation. The training and validation datasets are not shuffled between epochs during training. Namely, the validation dataset is not ‘seen’ by the DNN model so that we can validate the generalization ability from the training curves.

The input to the DNN models has a size of $256 \times 256 \times 4$ (4 Phase Approximants estimated from each interferogram), while the size of the ground truth RI maps is $256 \times 256 \times 100$. The original size of each interferogram is 1024×1024 ; however, the spectral filter we used for decoupling the 4 illumination angles cropped out the high-frequency information, as shown in Fig. 5, which does not have a significant influence since the phase maps do not contain too much detailed information. We down sampled each estimated Phase Approximants to a size of 256×256 to accelerate the computation of ground truth RI maps and the training of ML engines, without too much loss of the spatial resolution. The RI maps reconstructed by LT-BPM actually have a size of $256 \times 256 \times 256$. However, we only kept the central 100 layers for training the DNN models, since most cells have relatively flat structures, and the rest layers have little information of the cell’s structure. Removing these layers boosts the training speed and avoids artifacts due to noise from the medium. The reconstruction time for each RI map using LT-BPM is ~20 mins, while the average training time for DNN-L, DNN-H, and DNN-S is 6-7 hours when training on the dataset of ~900 NIH/3T3 cells. The inference time for our trained DNN model is 0.68 s, while the computational time for Phase Approximants is 1.5 sec, and the time used for linear fitting is 0.03 sec. The comparison of different ODT methods’ computation time is shown in Table S1, where we see SILACT outperforms other ODT methods according to computation time. The phase retrieval algorithm is universal that searches all the orders in the Fourier domain and performs linear filtering and shifting after all the orders are found. This algorithm can be accommodated to different optical system settings while taking more time. Besides, the DNN inference time could also be improved if we use the computer with better-performed GPUs.

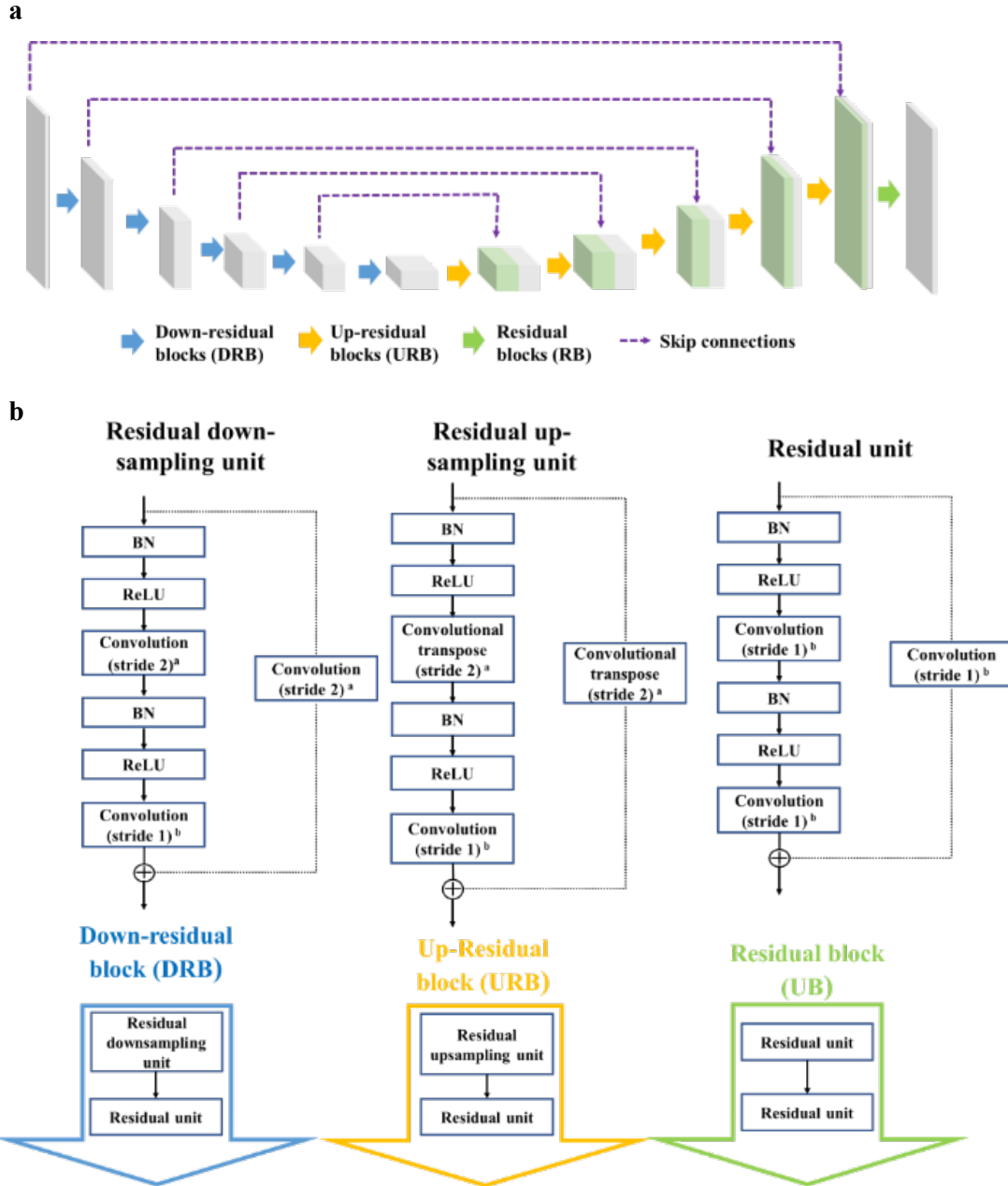


Figure S4. The basic architecture of machine learning engine. (a) The general structure for the U-Net with residual blocks. (b) Detailed structures for Down-residual Blocks (DRB), Up-residual Blocks (URB), and Residual Blocks (RB). Superscript a and b denote different kernel sizes and strides, listed as follows: a) Kernel size: (3, 3), strides: (2, 2). b) Kernel size: (3, 3), strides: (1, 1).

3D RI reconstruction method	Image acquisition time	Phase retrieval time	3D reconstruction time	Linear fitting time	Total computation time
ROWOR	3.90 msec	18.93 sec	19.98 sec	N/A	38.91 sec
ROWR	3.90 msec	18.93 sec	~7.8 min	N/A	~8.1 min

LT-BPM	3.90 msec	18.93 sec	~20 min	N/A	~20 min
SILACT	0.08 msec	1.50 sec	0.68 sec	0.03 sec	2.21 sec

Table S1. The comparison of computation time when using different 3D RI reconstruction methods. Different columns show the computation time of different reconstruction steps, while different rows are the computation time of different reconstruction methods.

6. Additional reconstructions with SILACT for further assessment

In Fig. S5, we show a further comparison of reconstructions by LT-BPM with the same four illumination angles as shown in Fig. S1a, SILACT without LS-DNN, SILACT with LS-DNN, and the ground truth (generated by LT-BPM with all 49 illumination angles). From the comparison, we can clearly see that SILACT method with LS-DNN enhancement enables reconstructions with optimal visualization performances for all classes of cells.

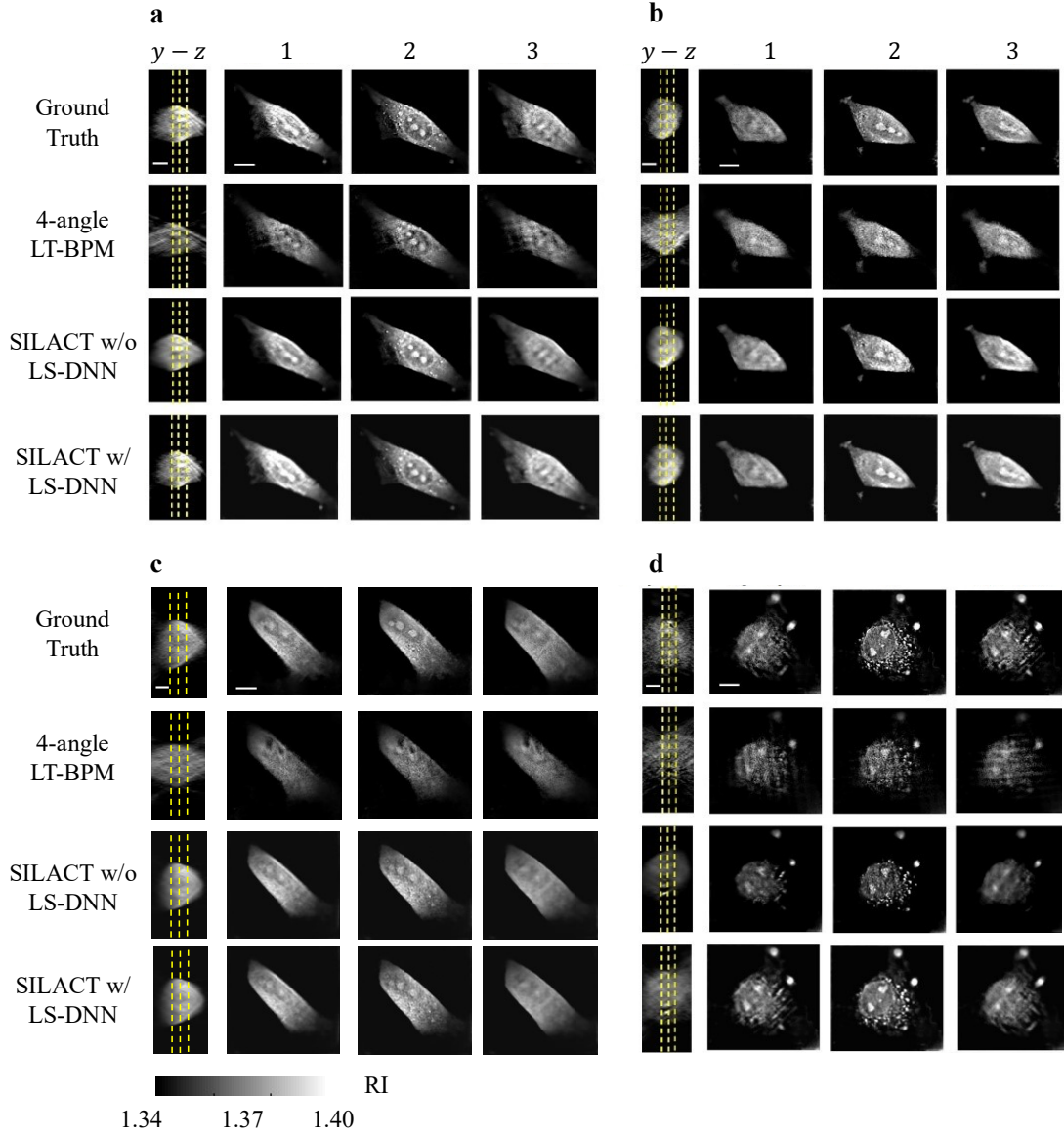


Figure S5. The comparison of visualized results with different reconstruction methods. The y-z cross-sections, and the x-y cross-sections at the z locations (1) $z_0 - 1\mu m$, (2) z_0 and (3) $z_0 + 1\mu m$ of (a) NIH/3T3 cells, (b) HEK293T cells (c) HeLa cells, and (d) COS-7 cells are shown, where z_0 indicates the central plane. The different z locations are also labelled with yellow dashed lines in y-z cross-sections. The scale bar in the x-y plane is $10\mu m$, and the scale bar along the z-axis is $2\mu m$.

7. Impact of different training loss metrics

In Fig. S6, we investigate the role of the training loss function in the quality of the reconstructions. From Fig. S6, we see that all other factors same, NPCC as the training loss function gives us the best reconstructions when incorporating linear fitting, in terms of all quantitative metrics (PCC, MAE, and RMSE). Therefore, the use of NPCC as the training loss function, despite the need for a further step of linear fitting, is completely justified.

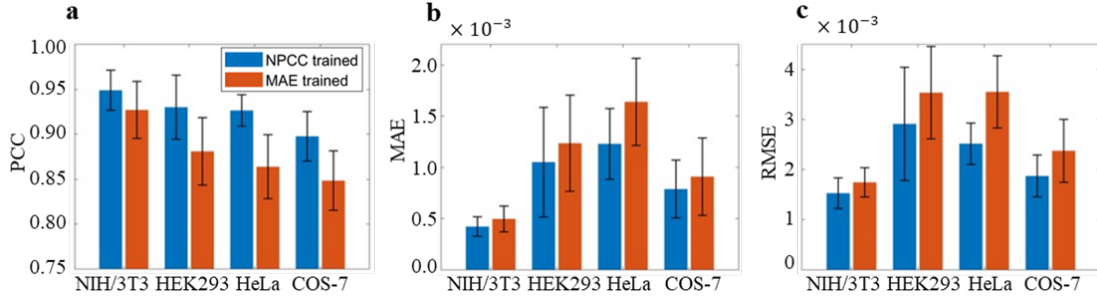


Figure S6. the quantitative evaluation of SILACT's reconstruction with different training loss metrics. (a)-(c), the PCC, MAE loss, and RMSE loss of the ground truth RI map (LT-BPM reconstruction with 49 angles) and the RI map reconstructed with our trained DNN on the dependence of different cell species, respectively. The DNN model used for producing the RI map is trained with a dataset of NIH/3T3 cells. The blue bars show the results of the DNN trained with NPCC loss function, while the red bars demonstrate the results of the MAE loss function trained DNN model.

8. Analysis of cells

In the analysis of red blood cells' deformation, we utilized the *regionprops3* function in MATLAB to calculate the eigenvalues of RBCs' shape in 3D. Firstly, we chose the first eigenvalue ε_1 , which is the length of the long axis of the fitted ellipsoid shape. Then we pick the last two eigenvalues the mean value of the last two eigenvalues, ε_2 and ε_3 , and choose the larger one as the dominator. Then the ratio of the lengths of long and short axes is calculated as the eccentricity ζ of the RBC, expressed as:

$$\zeta = \frac{\varepsilon_1}{\llbracket \max(\varepsilon)_2, \varepsilon_3 \rrbracket}. \quad (\text{S17})$$

For the analysis of NIH/3T3 cells in the microfluidic channel, we also used the *regionprops3* function in MATLAB to calculate the volumes and surface areas in 3D. The dry mass of the cell m could be calculated with the formula shown below, according to ref (8).

$$m = \frac{\bar{\chi}V}{2n_m\alpha}, \quad (\text{S18})$$

where V is the calculated volume of the cell; n_m is the refractive index of the medium, which is 1.337 here since we used the PBS to culture the cells; α is the refractive index increment, which equals $0.2 \text{ g}\cdot\text{mL}^{-1}$; $\bar{\chi}$ is the mean scattering potential in the cell region, and the scattering potential χ at position \mathbf{r} can be calculated as:

$$\chi(\mathbf{r}) = n_c^2(\mathbf{r}) - n_m^2, \quad (\text{S19})$$

where $n_c(\mathbf{r})$ is the refractive index in the cell at position \mathbf{r} .

For determining the throughput of the 3D flow image cytometry based on our SILACT method, we utilized a customized algorithm which is realized on the platform of MATLAB. Since most of the cells are flowing near the focal plane of our interferometric microscope, and usually we only have one layer of cells, we take the

centre layer (the 50th layer) of the reconstructed RI map for the cell segmentation and counting. We set a threshold RI value (usually 1.37) for extracting the cell regions, and then use *bwlabel* function in MATLAB to segment different cells. To avoid the repeated counting of the cells, we calculated the Pearson Correlation Coefficient (PCC) values between the cells in two adjacent frames and exclude the cells which have PCC values larger than 0.90 with cells in previous frames. In our analysis, we counted 936 NIH/3T3 cells and 1340 RBCs in 2000 frames under a frame rate of 12,500 fps. Namely, we can reach a throughput of analysing the 3D RI maps of over 5,000 cells (5850 NIH/3T3 cells or 8500 RBCs) in one second with our flow image cytometer based on SILACT.

9. NIH/3T3 cell trained vs. RBC trained models when imaging deformed RBCs in microfluidic channels

In Section 2.5 of the main text, we demonstrate the application of SILACT to reconstruct and analyse the 3D morphology of deformed RBCs when flowing through a microchannel. In this section, we compare the prediction results on RBCs of two ML engines (models), one is trained with NIH/3T3 cells, as we discussed in the main text, while the other is trained with RBCs (500 image pairs for training, 5% for validation), in order to justify our adoption of ML engine trained with RBCs and further demonstrate the generalization ability of our SILACT method.

Firstly, the ML engines trained with NIH/3T3 cells and RBCs are tested on 100 images of RBCs in static states, and their testing results are quantitatively evaluated with MAE and PCC and shown in Fig. S7b-c. It seems that the 3T3 cell trained ML engine has worse performances, however, its averaged testing PCC value is still larger than 0.85 (0.89 ± 0.05). On the other hand, the testing PCC value of RBC trained ML engine is 0.94 ± 0.03 .

After that, we used these two ML engines to predict the 3D RI maps of the deformed flowing RBCs that we showed in Fig. 4. We picked the same single RBC in Fig. 4c and showed its 3D RI maps reconstructed by ML engines trained with 3T3 cells and RBCs at different time points in Fig. S7a. We can see despite some mismatches in fine structures, the 3D contours of the RBCs are consistent in both models. Furthermore, the extracted morphological indices (3D eccentricity & volume) are quite similar in the results reconstructed by both ML engines, as shown in Fig. S7 d-e.

In the main text, we showed the results predicted by the ML engine trained with RBCs to avoid the emergence of any controversial fine structure generated by the DNN model trained with NIH/3T3 cells. However, it is worth noticing that we can extract similar morphological parameters no matter which ML engine the results are predicted from, which further validate the outstanding generalization ability of SILACT.

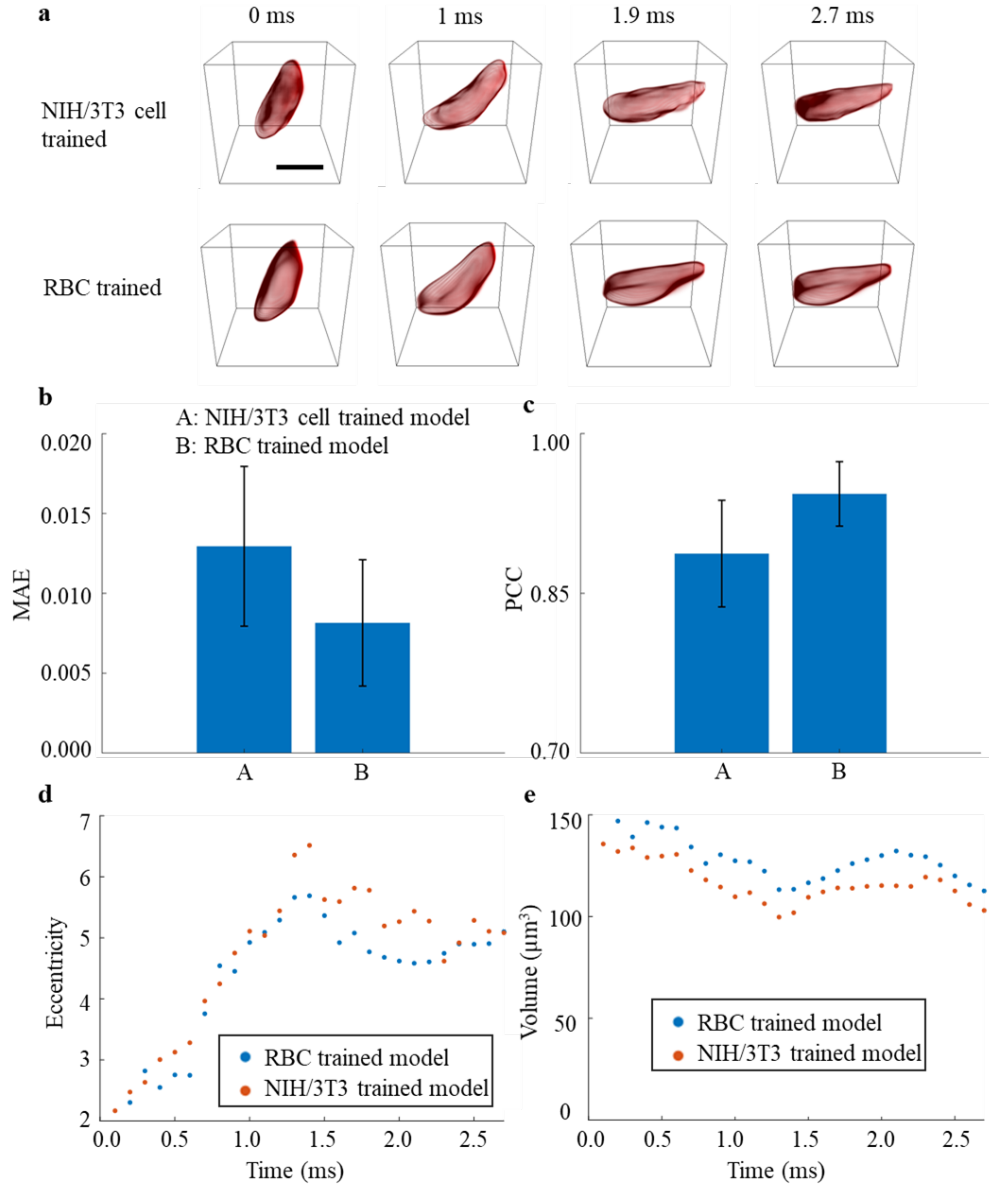


Figure S7. The comparison of performances of the NIH/3T3 cell trained model and RBC trained model on the deformed RBCs in the microfluidic channel underflow. (a) The 3D rendering of the selected cell's RI maps predicted by the NIH/3T3 cell trained model and the RBC trained model at 0, 1, 1.9, and 2.7 ms. The scale bar here denotes 10 μm . (b) and (c), the MAE and PCC between the ground truth and prediction results on tested RBC cells with NIH/3T3 cell trained model and RBC trained model. (d) and (e), the comparison of the eccentricity and volume over time of the deformed RBC when predicted with NIH/3T3 cell trained and RBC trained models.

10. Microfluidic chip fabrication

The microfluidic device used to image cells under the proposed SILACT was fabricated following standard soft lithography (9, 10). Positive photoresist AZ 4620 (AZ Electronic Materials) was spun on a 4" silicon wafer at a speed of 5000 RPM for 30 s to obtain a master mold of 5 μm height. The photoresist was exposed under UV (ABM-USA) through a premade Cr/Au mask (Microcad Photo-Mask Ltd) carrying the designed microchannel. The exposed photoresist was then developed by a developer

(AZ400K: deionized water = 1:4, by volume). On the other hand, to obtain a master mold of 18 μm and 30 μm height, negative photoresist SU8-3050 (Kayaku Advanced Materials) was spun on a 4" silicon wafer at the speed of 5000 RPM and 3000 RPM for 30 s, respectively. The SU-8 patterns were developed and baked following the manufacturer's protocol. The developed patterns (for both AZ 4620 and SU8-3050) were subsequently hard backed at 180 °C for 3 hr with a ramp of 1 °C/min on a programmable hot plate (Torrey Pines Scientific). The surface of produced mold was treated with a thin anti-adhesive hydrophobic layer of Trichloro(1H,1H,2H,2H-perfluorooctyl) silane (TOFC, Sigma Aldrich, catalog no. 448931) by placing the mold together with TOFC in a desiccator for 30 min under a fume hood (10) (Fig. S8 a-d).

Subsequently, pre-mixed polydimethylsiloxane (PDMS) (SYLGARD™ 184 Silicone Elastomer Kit, base: curing agent = 1:10) was spun on fabricated master mold at 500 RPM for 30 s to produce a ~ 200 μm layer of PDMS film (11). The PDMS film was cured at 70 °C for 2 hours (Fig. S8e). To create the inlet and outlet of the microchannel, firstly, two pre-made PDMS blocks (20 \times 20 \times 3 mm) and the surface of the cured PDMS film were activated by oxygen plasma (Harrick Plasma, Expanded Plasma Cleaner).

The PDMS blocks were then assembled onto the PDMS film by aligning on top of the inlet and outlet regions of the microchannel under an inverted microscope (Fig. S8f). The master mold along with the PDMS film was baked at 110 °C for 20 min to ensure permanent bonding of the assembly. Finally, the PDMS assembly was carefully peeled off, hole punched, and sealed with a No. 1 cover glass by plasma bonding (Fig. S8g-i). The thickness of the microfluidic chip was measured with a digital Vernier caliper (Fig. S9a) for a thickness below 350 μm , to satisfy the required working distance of objectives OL1 and OL2. The fabricated microfluidic device was then placed at the observation stage of the proposed ODT for subsequent measurements (Fig. S9b). The microfluidic chips and tubing accessories were sterilized before each experiment. The samples were introduced into the microfluidic device by a syringe pump (Harvard Apparatus, PHD 2000) at the designated volumetric flow rate.

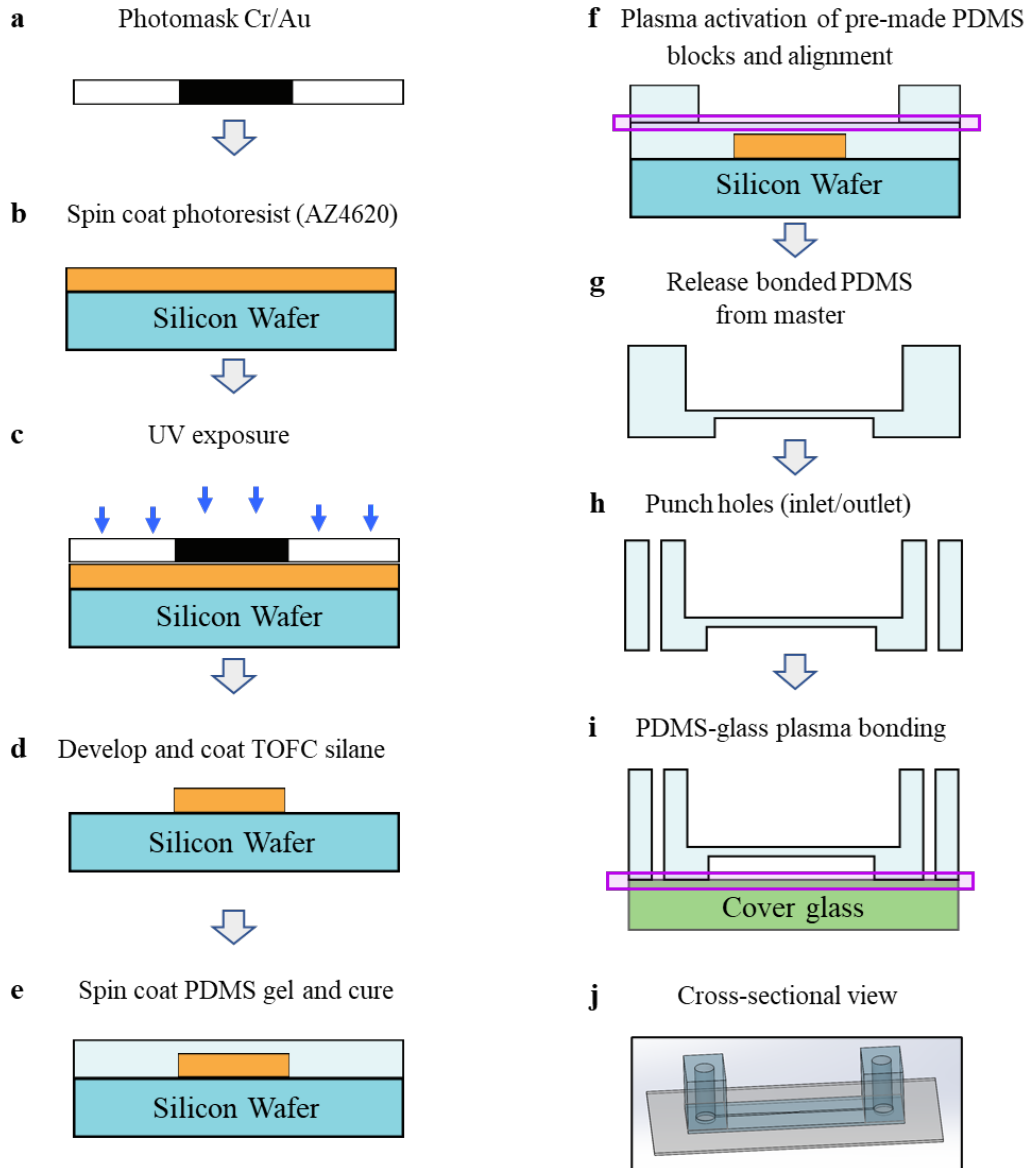


Figure S8: Fabrication procedures of the PDMS microfluidic chip. (a) Design of the microfluidic channel on Cr/Au photomask. (b) Spin coat positive photoresist (AZ 4620 or SU8-3050) on a 4" silicon wafer. (c) Expose the coated photoresist under UV to through the photomask to transfer the pattern on the wafer. (d) Develop the exposed photoresist and treat the surface of the master mold with TOFC silane. (e) Spin coat PDMS gel on the master mold and cure. (f) Place pre-made PDMS block on the inlet and outlet region of the thin PDMS film and bond with plasma activation. (g-i) Release the PDMS from the master mold, punch inlet, and outlet holes, and finally seal with a No. 1 cover glass. (j) Showed the overview of the PDMS microfluidic chip.

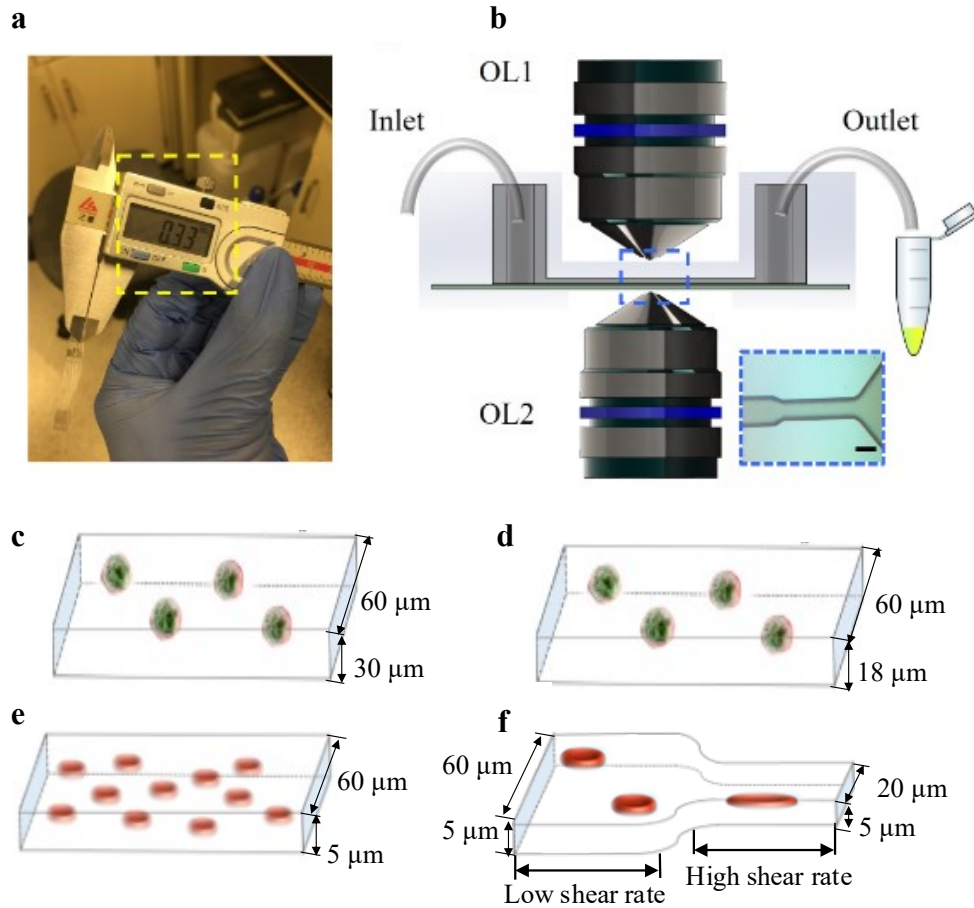


Figure S9: Assembling of the microfluidic chip with ODT. (a) Thickness of the prepared microfluidic chip to ensure the placement in between the two objective lenses (b) on the proposed ODT system. (c)-(e), The microfluidic channels for NIH/3T3 cells, Jurkat T cells and Mouse RBCs, the widths and heights are shown in each subfigure. (f) The design of the microfluidic channel. The left region is wider, where the shear rate of the cell is lower, while the right region is narrower with a higher shear rate.

11. Preparation of measurement buffer (MB)

A solution of 15 mPa.s viscosity was prepared as the measurement buffer to avoid cell sedimentation and to generate a sufficient shear rate ($\sim 20,000 \text{ s}^{-1}$) at low cell velocity (12). The measurement buffer was prepared by dissolving methylcellulose in a physiological buffer (PBS) (0.5%, w/v). To fully dissolve the methylcellulose (13), the solution was mixed with a rotary mixture for 24 hours. The undissolved residue was removed by a vacuum filter (MF-Millipore™ Membrane Filter, 0.22 μm pore size). The viscosity of the obtained buffer was characterized (Fig. S10) using a rotational rheometer (Malvern Kinexus Lab+) and calibrated to 15 mPa.s by the addition of PBS. Inset of Fig. S10 showed a negligible portion of cells were sedimented in MB compared to PBS after 1 hr of incubation at 37°C. Therefore, to enable a prolonged duration of the cell observations, red blood cells were suspended in the MB for the subsequent experiments.

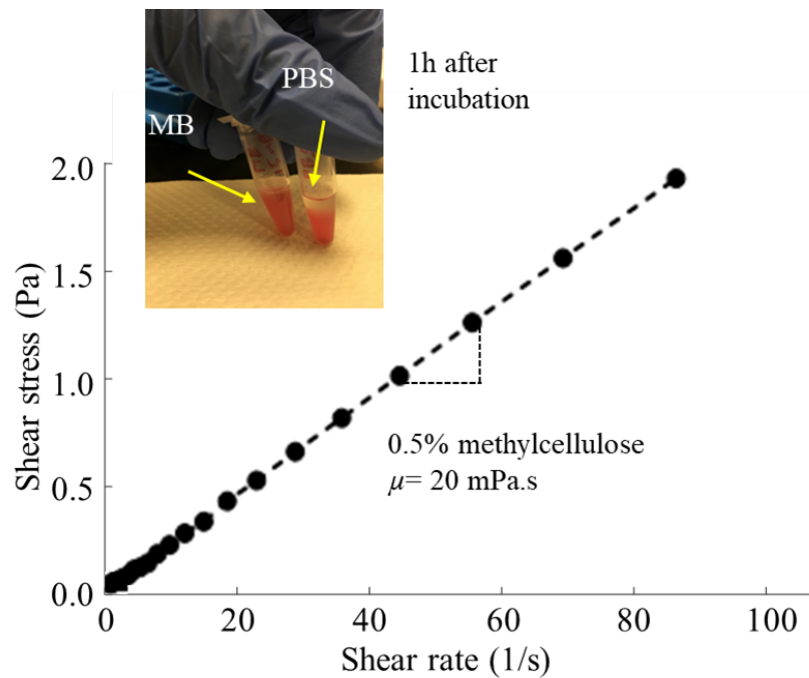


Figure S10: Shear rate to the stress of the prepared measurement buffer. The viscosity was adjusted to 15 mPa.s by adding PBS to prepare RBCs suspension for subsequent experiments. Inset showed a negligible portion of cells were sedimented in MB after 1h of incubation in 37°C.

References:

1. Y. Sung *et al.*, Optical diffraction tomography for high resolution live cell imaging. *Opt. Express* **17**, 1977–1979 (2009). doi: 10.1364/OE.17.000266.
2. J. Lim *et al.*, Comparative study of iterative reconstruction algorithms for missing cone problems in optical diffraction tomography. *Opt. Express* **23**, 16933–16948 (2015). doi: 10.1364/OE.23.016933.
3. U. S. Kamilov *et al.*, Learning approach to optical tomography. *Optica* **2**, 517–522 (2015). doi: 10.1364/OPTICA.2.000517.
4. U. S. Kamilov *et al.*, Optical Tomographic Image Reconstruction Based on Beam Propagation and Sparse Regularization. *IEEE Trans. Comput. Imaging* **2**, 59–70 (2016). doi: 10.1109/TCI.2016.2519261.
5. R. R. Dasari, Y. Sung, Deterministic regularization of three-dimensional optical diffraction tomography. *JOSA A* **28**, 1554–1561 (2011). doi: 10.1364/JOSAA.28.001554.
6. M. Ziemczonok, A. Kuś, P. Wasylczyk, M. Kujawińska, 3D-printed biological cell phantom for testing 3D quantitative phase imaging systems. *Sci. Rep.* **9**, 18872 (2019). doi: 10.1038/s41598-019-55330-4.
7. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016).
8. T. H. Nguyen, M. E. Kandel, M. Rubessa, M. B. Wheeler, G. Popescu, Gradient

- light interference microscopy for 3D imaging of unlabeled specimens. *Nat. Commun.* **8**, 210 (2017). doi:10.1038/s41467-017-00190-7.
9. X. Cui *et al.*, Lensless high-resolution on-chip optofluidic microscopes for *Caenorhabditis elegans* and cell imaging. *Proc. Natl. Acad. Sci.* **105**, 10670–10675 (2008). doi: 10.1073/pnas.0804612105.
 10. D. Qin, Y. Xia, G. M. Whitesides, Soft lithography for micro- and nanoscale patterning. *Nat. Protoc.* **5**, 491–502 (2010). doi: 10.1038/nprot.2009.234.
 11. W. Y. Zhang, G. S. Ferguson, Elastomer-supported cold welding for room temperature wafer-level bonding. *Proc. IEEE Micr. Elect.* 741–744 (2004). doi: 10.1109/MEMS.2004.1290691.
 12. E. N. L. R. Byron Bird, Warren E. Stewart, *Transport Phenomena* (John Wiley and Sons, New York, 2007).
 13. M. Urbanska, P. Rosendahl, M. Kräter, J. Guck, High-throughput single-cell mechanical phenotyping with real-time deformability cytometry. *Methods Cell Biol.* **147**, 175–198 (2018). doi: 10.1016/bs.mcb.2018.06.009.