Title

Deep Learning in Optical Microscopy, Holographic Imaging and Sensing

Permalink

https://escholarship.org/uc/item/8k03d1g2

Author

Liu, Tairan

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Deep Learning in Optical Microscopy, Holographic Imaging and Sensing

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Electrical and Computer Engineering

by

Tairan Liu

2022

ABSTRACT OF THE DISSERTATION

Deep Learning in Optical Microscopy, Holographic Imaging and Sensing

by

Tairan Liu

Doctor of Philosophy in Electrical and Computer Engineering

University of California, Los Angeles, 2022

Professor Aydogan Ozcan, Chair

The microscopy imaging technique has been employed as the gold-standard method for diagnosing numerous diseases for hundreds of years. However, the dependence on high-end optical components of traditional optical microscopes may often limit their usage in many applications. Recent developments in deep learning-enabled computational imaging techniques have revolutionized the field achieving both faster speed and higher image quality while maintaining the simplicity of the optical system.

In the first part of this dissertation, a set of novel deep learning-enabled microscopy imaging techniques is introduced to perform super-resolution, color holography, and quantitative polarization imaging, which aims for improving the performance of the existing optical system. Firstly, deep learning was adopted to enhance the resolution of both pixel size-limited and diffraction-limited coherent imaging systems, providing a rapid, non-iterative method to improve the space-bandwidth product of coherent imaging systems. Then, a high-fidelity color image reconstruction method using a single hologram is presented, where deep learning was used to simultaneously eliminate the missing-phase-related artifacts and correct the color distortion. In addition, another deep learning-enabled holographic polarization microscope is demonstrated, which can obtain quantitative birefringence retardance and orientation information of the specimen from a phase-recovered hologram from one polarizer/analyzer pair.

In the second part of this dissertation, deep learning is further applied to various biological imaging or sensing applications enabling these systems to perform virtual histology staining, cell classification, and pathogen detection. A digital staining technique is first demonstrated to transform the quantitative phase images (QPI) of label-free tissue sections into images equivalent to the brightfield microscopy images of the same tissue sections that are histologically stained. Next, using time-lapse lensless speckle imaging and a deep learning classifier, a computational cytometer is shown to rapidly detect magnetic bead-conjugated rare cells of interest in three dimensions (3D). Lastly, two deep learning-based pathogen detection frameworks are presented. A bacterial colony-forming-unit (CFU) detection system exploiting a thin-film-transistor (TFT)-based image sensor array is firstly shown which can save ~12 hours compared to the Environmental Protection Agency (EPA)-approved methods. Then, a stain-free quantitative viral plaque assay framework is presented which could automatically detect the first cell lysing events due to the viral replication as early as 5 hours after the incubation and achieved a >90% detection rate for the plaque-forming units (PFUs) with 100% specificity in <20 hours, providing major time savings compared to the traditional plaque assays that take ≥48 hours.

The dissertation of Tairan Liu is approved.

Mona Jarrahi

Dino Di Carlo

Achuta Kadambi

John D. Fitzgerald

Aydogan Ozcan, Committee Chair

University of California, Los Angeles

2022

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my Ph.D. advisor, Prof. Aydogan Ozcan, to the greatest extent for supporting me throughout my Ph.D. study and research. I feel extremely lucky to have such a passionate, creative, rational, and caring professor as my Ph.D. advisor. Along this journey with prof. Ozcan, I have learned not only about technical skills and research knowledge, but also how to build, communicate, and lead a team among many other human skills. This experience is of no doubt the most invaluable treasure for my future career.

I would also like to thank my doctoral committee members, Prof. Mona Jarrahi, Prof. Dino Di Carlo, Prof. Achuta Kadambi, Dr. John FitzGerald for their invaluable scientific advice and career suggestions.

In addition, great colleagues and collaborators are of exceptional importance during my Ph.D. study. Here, I would like to first give my special thanks to Yibo Zhang and Yair Rivenson for their guidance upon my arrival at UCLA. I would also like to thank Kevin De Haan, Hatice Ceylan Koydemir, Yuzhu Li, Hongda Wang, Yi Luo, Yichen Wu, Bijie Bai, Jingxi Li, Yijie Zhang, Luzhe Huang, Xilin Yang, Hanlong Chen, Zhensong Wei, Yilin Luo, Zoltán Göröcs, Hyouarm Joung, Merve Eryilmaz, Artem Goncharov, Zach Ballard, Calvin Brown, Aniruddha Ray, Derek Tseng, Mengxing Ouyang, Ethan Yang, Xin Zeng, David Karalli, Hongxiang Fu, Keelan O'Riordan, among many others for the support during my Ph.D. research.

Finally, I would like to express my sincere gratitude to my family for their unconditional love, support, and understanding throughout this adventure of my Ph.D. study.

Vita

Tairan Liu received his Bachelor of Engineering degree from the Beijing Institute of Technology, Beijing, China in 2015. Then he received a master's degree in Electrical and Computer Engineering from the University of Michigan, Ann Arbor, the USA in 2017. He started pursuing his Ph.D. in Electrical and Computer Engineering at the University of California, Los Angeles, USA. His expertise is focused on the field of biomedical imaging, computational imaging, digital holography, and signal and image processing.

Selected publications

Liu, T., Li, Y., Koydemir, H.C., Zhang, Y., Yang, E., Wang, H., Li, J., Bai, B. and Ozcan, A., (2022). Stain-free, rapid, and quantitative viral plaque assay using deep learning and holography. arXiv preprint arXiv:2207.00089.

Liu, T., De Haan, K., Rivenson, Y., Wei, Z., Zeng, X., Zhang, Y., & Ozcan, A. (2019). Deep learning-based super-resolution in coherent imaging systems. Scientific reports, 9(1), 1-13.

Liu, T., Wei, Z., Rivenson, Y., de Haan, K., Zhang, Y., Wu, Y., & Ozcan, A. (2019). Deep learning-based color holographic microscopy. Journal of biophotonics, 12(11), e201900107.

Liu, T., de Haan, K., Bai, B., Rivenson, Y., Luo, Y., Wang, H., ... & Ozcan, A. (2020). Deep learning-based holographic polarization microscopy. ACS photonics, 7(11), pp.3023-3034.

Li, Y., **Liu, T.**, Koydemir, H.C., Wang, H., O'Riordan, K., Bai, B., Haga, Y., Kobashi, J., Tanaka, H., Tamaru, T., Yamaguchi K., Ozcan, A. (2022) Deep Learning-enabled Detection and Classification of Bacterial Colonies using a Thin Film Transistor (TFT) Image Sensor. ACS Photonics, 9(7), pp.2455–2466

Rivenson, Y., **Liu, T.**, Wei, Z., Zhang, Y., de Haan, K., & Ozcan, A. (2019). PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. Light: Science & Applications, 8(1), 1-11.

Zhang, Y., Ouyang, M., Ray, A., **Liu, T.**, Kong, J., Bai, B., ... & Tsai, K. (2019). Computational cytometer based on magnetically modulated coherent imaging and deep learning. Light: Science & Applications, 8(1), 1-15.

Chapter 1 . Introduction to deep learning-enabled computational microscopy

The optical microscopy imaging technique has been employed as the gold-standard method for diagnosing numerous diseases for hundreds of years, leading to continuous advances and discoveries. However, the traditional lens-based optical microscope is limited by many drawbacks, including high system cost, limited throughput, bulky optical design, requirements for accurate alignment, etc. The development of the lensless imaging technique has largely overcome such burdens and offered many flexibilities in areas such as point-of-care diagnostics, lab-on-a-chip applications, global health applications, telemedicine, etc. Nevertheless, without performing many image quality enhancement techniques, such as phase recovery, pixel-super-resolution, hyper-spectral imaging, etc., it is often hard to obtain high image quality using such systems.

Recently, deep learning has emerged as a highly effective technique for solving inverse problems in microscopy. It has been applied to traditional inverse problems such as holographic image reconstruction, reconstruction of color images, super-resolution, quantitative polarization microscopy, as well as to perform cross-modality image transformations such as virtual labeling of histological tissue, live cells monitoring, and pathogen detection. In this dissertation, I will focus on discussing the above innovations enabled by deep learning. The dissertation is structured as follows:

Chapters 2-4 will focus on three different deep learning-enabled frameworks that improves the traditional lensfree microscopy technique. In Chapter 2, I will use lensfree microscopy as an example to discuss a central challenge for all microscopy techniques, i.e., spatial resolution enhancement, and demonstrate the success of the deep learning framework to be able to improve the spatial resolution in both pixel-size-limited case and diffraction-limited case. In

Chapter 3, I will demonstrate a deep learning-based color holography framework, where the network is used to simultaneously remove the missing-phase-related artifact and correct the color distortion when using only three illuminating wavelengths to form a color image in the lensfree imaging system. In Chapter 4, I will present a quantitative holographic polarization microscopy technique, where the system only requires one polarization state in its light path to reconstruct both the retardance and orientation information of a birefringent object.

Then, in Chapters 5-8, I will cover four biomedical applications of deep learning-based computational microscopy and the related techniques. In Chapter 5, a virtual staining technique will be introduced where a deep learning algorithm will be used to digitally convert the quantitative phase images of a label-free tissue into a bright-field image as if the sample is stained with histochemical stains. In Chapter 6, a compact lensfree cytometer will be presented, which is able to rapidly detect magnetic beads conjugated rare cells using time-lapse imaging and deep learning. In Chapter 7&8, two pathogen detection frameworks will be shown to automatically detect colony forming units of bacteria during their growth on a chromogenic agar plate and to detect plaque forming units of viruses during their growth in a stain-free manner.

Finally, I will summarize the dissertation in Chapter 9.

Chapter 2 . Enhancing resolution for coherent imaging systems with deep learning

2.1 Introduction

Coherent imaging systems have many advantages for applications where the specimen's complex field information is of interest [1]. Since Gabor's seminal work, various optical and numerical techniques have been suggested [2] to acquire the complex field of a coherently illuminated specimen. This has allowed for the characterization of absorption and scattering properties of a sample, as well as enabling numerical refocusing at different depths within that sample volume. To infer an object's complex field in a coherent optical imaging system, the "missing phase" needs be retrieved. A classical solution to this missing phase problem is given by off-axis holography [3,4], which in general results in a reduction of the space-bandwidth product of the imaging system. In-line holographic imaging, which can be used to design compact microscopes [5], has utilized measurement diversity to generate a set of physical constraints for iterative phase retrieval [6–10]. Recently, deep-learning based holographic image reconstruction techniques have also been demonstrated to create a high-fidelity reconstruction from a single in-line hologram [11–13], and are capable of further extending the depth-of-field of the reconstructed image [14].

Several approaches have been demonstrated to improve the resolution of coherent imaging systems [15–20]. Most of these techniques require sequential measurements and assume that the object is quasi-static while a diverse set of measurements are performed on it. These measurements often require the use of additional hardware or sacrifice a degree of freedom such as the sample field-of-view [21]. In recent years, sparsity-based holographic reconstruction methods have also demonstrated that they are capable of increasing the resolution of coherent imaging systems without the need for additional measurements or hardware [22–25]. Sparse signal recovery methods employed in coherent imaging are based on iterative optimization algorithms. These methods usually involve a comprehensive search over a parameter space to obtain the optimal object image and generally result in longer reconstruction times.

Deep learning-based approaches for super-resolution of incoherent microscopy modalities such as brightfield and fluorescence microscopy have also recently emerged [26–30]. These data-driven super-resolution approaches produce a trained deep convolutional neural network that learns to transform low-resolution images into high-resolution images in a single feed-forward (i.e., non-iterative) step. Generative adversarial networks (GANs) [31] are a form of deep neural network training framework that can be used to ensure that the generated image is sharp and realistic. A GAN is made up of two separate networks. A generator network is used to generate an image that has the same features as the label (ground truth) image, and a discriminator network tries to distinguish between the generated and label (ground truth) images.

Here, we apply deep learning to enhance the resolution of coherent imaging systems and demonstrate a conditional GAN that is trained to super-resolve both *pixel-limited* and *diffraction-limited* images. Furthermore, we demonstrate the success of this framework on biomedical samples such as thin sections of lung tissue and Papanicolaou (Pap) smear samples. We quantify our results using the structural similarity index (SSIM) [32] and spatial frequency content of the network's output images in comparison to the higher resolution images (which constitute our ground truth). This data-driven image super-resolution framework is applicable to enhance the performance of various coherent imaging systems.

Part of this chapter has been published in:

Liu, T., De Haan, K., Rivenson, Y., Wei, Z., Zeng, X., Zhang, Y., & Ozcan, A. (2019). Deep learning-based super-resolution in coherent imaging systems. Scientific reports, 9(1), 1-13.

2.2 Materials and methods

First we briefly summarize the methods that we have used; sub-sequent subsections will provide more information on specific methods employed in our work. We applied the presented deep learning-based super-resolution approach to two separate in-line holographic imaging geometries to demonstrate the efficacy of the technique. As illustrated in Fig. 2.1a and Fig. 2.1b, the two implemented configurations were a pixel size-limited system (to be referred to as System A) and a diffraction-limited coherent microscopy system (to be referred to as System B). Despite using different methods to create the super-resolved images, as a result of the different image formation models, both of these systems followed similar general hologram reconstruction steps, i.e., 1. Raw holograms were collected at different sample to sensor distances, 2. Autofocus was used to determine the accurate sample to sensor distances, 3. Phase was recovered using a multi-height phase recovery algorithm. These steps will be detailed in the following subsections of the Methods [5,7,33–36].

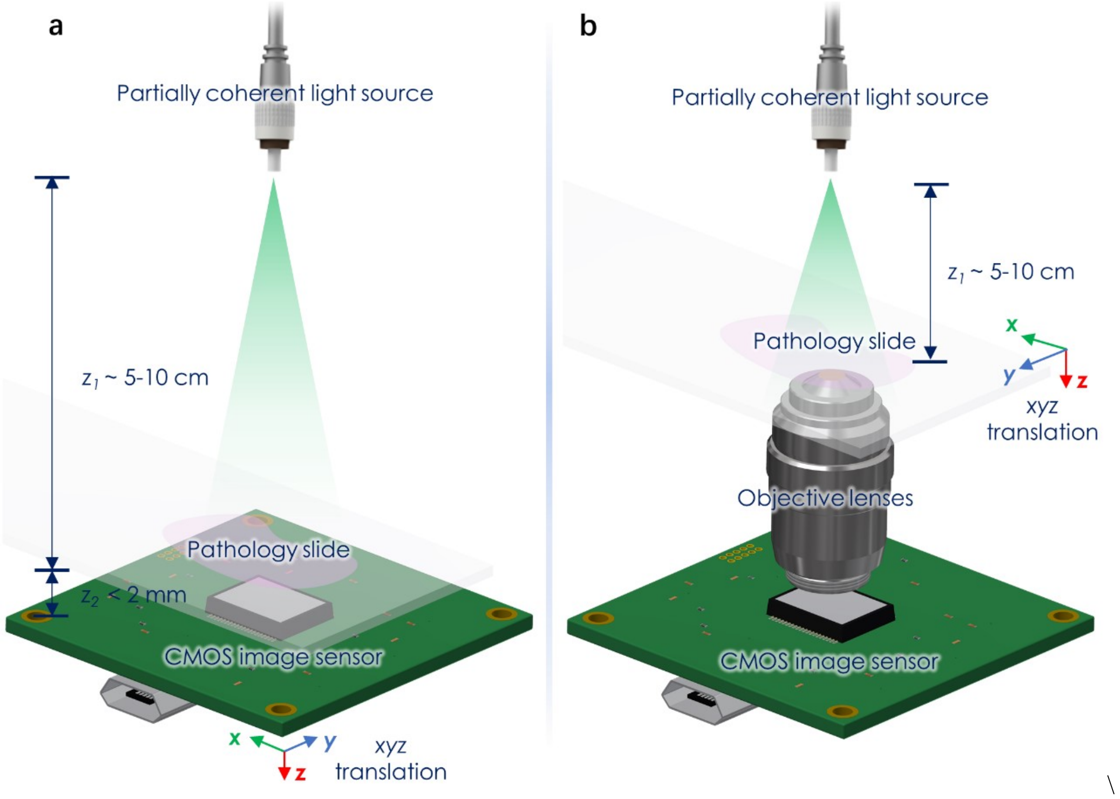


Figure 2.1 Schematic of the coherent imaging systems. (a) A Lens-free on-chip holographic microscope. The sample is placed at a short distance (z2 < 2 mm) above the image sensor chip. The resolution of this lensless on-chip imaging modality (without the use of additional degrees of freedom) is pixel size-limited due to its unit magnification. (b) A lens-based in-line holographic microscope, implemented by removing the condenser and switching the illumination to a partially-coherent light source on a conventional bright-field microscope. The resolution in this case is limited by the NA of the objective lens.

For the pixel-super-resolution network (System A), the network training process is demonstrated in Fig. 2.2, which summarizes both the hologram reconstruction procedure as well as the image super-resolving technique with and without using the network. The real and imaginary components of the phase recovered image pairs were used to train the network.

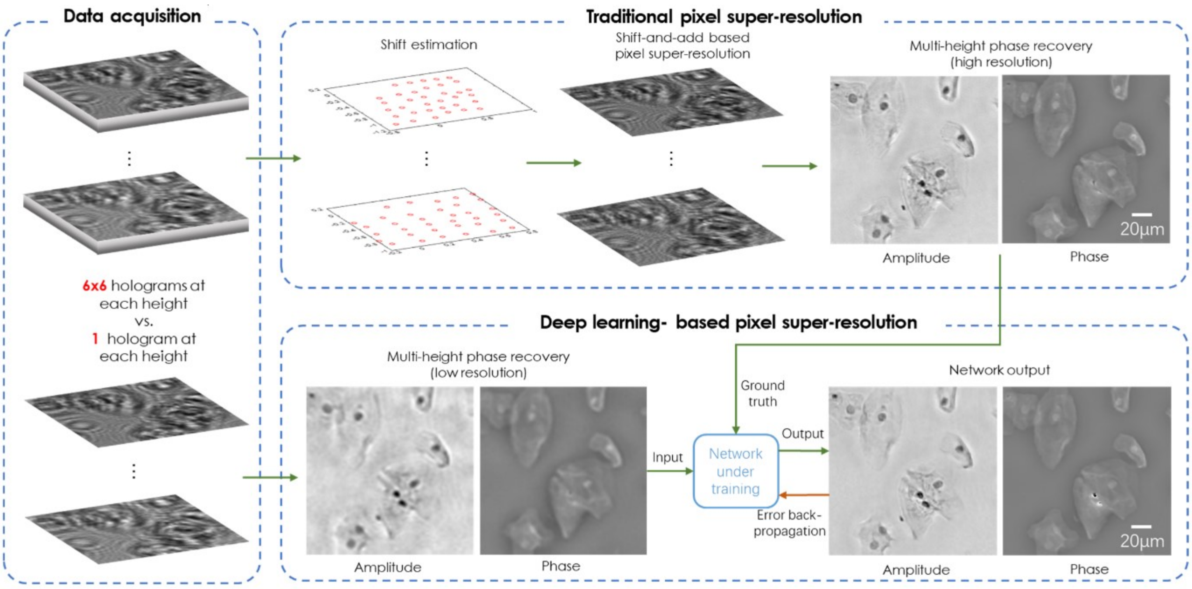


Figure 2.2 Schematic of the training process for deep-learning based pixel super-resolution.

For the diffraction-limited super-resolution network (System B), the network training process was demonstrated in Fig. 2.3. In this case only the phase channel was used to train the network.

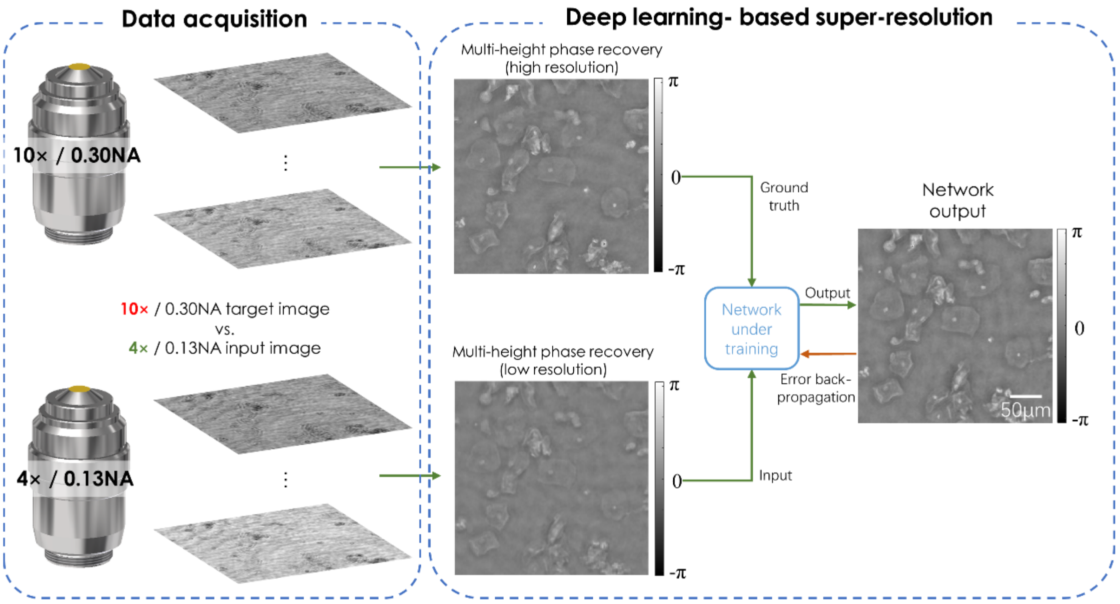


Figure 2.3 Schematic of the training process for deep learning-based optical super-resolution for an NA-limited coherent imaging system.

Generation of network input and ground truth super-resolved image labels.

For the pixel size-limited coherent imaging system (System A), the super-resolved images were created by collecting multiple low-resolution holograms at different lateral positions, where the CMOS image sensor was sub-pixel shifted by a mechanical stage to create a shift table. Once an accurate shift table was estimated, a shift-and-add based pixel super-resolution algorithm [33] was applied. The set-up used an illumination wavelength of 550 nm with a bandwidth ($\Delta\lambda$) of ~2 nm (WhiteLase Micro with acousto-optic tunable filter, NKT Photonics), a single mode fiber (QPMJ-3S2.5A-488-3.5/125-1-0.3-1) with a core diameter of ~3.5 µm and a source-to-sample distance (z_1) of ~5 cm. As a result, the effective spatial coherence diameter at the sensor plane was larger than the width of the CMOS imager chip used in our on-chip imaging system. Therefore, the achievable resolution is limited by the temporal coherence length of the illumination [37], which is defined as:

$$\Delta L_c \approx \sqrt{\frac{2\ln 2}{\pi}} \cdot \frac{\lambda^2}{n\Delta\lambda} = 100.47\ \mu\text{m} \tag{1}$$

where n=1 is the refractive index. Assuming a sample-to-sensor distance (z_2) of ~300 µm, the effective numerical aperture (NA) of the set-up was limited by the temporal coherence of the source, and is estimated to be:

$$\text{NA} = n\sin\theta = n\sqrt{1-\cos^2\theta} = n\sqrt{1-\left(\frac{z_2}{z_2+\Delta L_c}\right)^2} \approx 0.6624 \tag{2}$$

Based on this effective numerical aperture and ignoring the pixel size at the hologram plane, the achievable coherence-limited resolution of our on-chip microscope is approximated as [4]:

$$d \propto \frac{\lambda}{\text{NA}} = \frac{0.55}{0.6624} = 0.8303\ \mu\text{m} \tag{3}$$

At the hologram/detector plane, however, the effective pixel pitch of the CMOS image sensor (IMX 081, Sony RGB sensor, pixel size of 1.12 µm) using only one color channel is 2.24 µm. Based on this, the effective pixel size for each super-resolved image after the application of the pixel super-resolution algorithm to 4 raw holograms (2×2 lateral positions), 9 raw holograms (3×3 lateral positions), and 36 raw holograms (6×6 lateral positions) are 1.12 μm, 0.7467 μm and 0.3733 μm, respectively. Based on Equation (3), the effective pixel size achieved by pixel super-resolution using 6×6 lateral positions can adequately sample the specimen's holographic diffraction pattern and is limited by temporal coherence. All of the other images (using 1×1, 2×2 and 3×3 raw holograms) remain pixel-limited in their achievable spatial resolution. This pixel-limited resolution of an on-chip holographic microscope is a result of its unit magnification. This allows the imaging system to have a large imaging field-of-view (FOV) that is only limited by the active area of the opto-electronic image sensor chip. This can easily reach 20-30 mm^2 and >10 cm^2 using state-of-the-art CMOS and CCD imagers, respectively [5].

For the second set-up (System B), which used lens-based holographic imaging for diffraction-limited coherent microscopy, the low- and high-resolution images were acquired with different objective lenses. For this set-up, the illumination was performed using a fiber coupled laser diode with an illumination wavelength of 532 nm. A 4×/0.13NA objective lens was used to acquire lower resolution images, achieving a diffraction limited resolution of ~4.09 μm and an effective pixel size of ~1.625 μm. A 10×/0.30NA objective lens was used to acquire the higher resolution images (ground truth labels), achieving a resolution of 1.773 μm and an effective pixel size of ~0.65 μm.

Autofocusing and singular value decomposition-based background subtraction.

For both types of coherent imaging systems, holograms at 8 different sample-to-sensor distances were collected to perform the multi-height phase recovery [5,7,33–36]. This algorithm requires accurate knowledge of the sample-to-sensor distances used. These were estimated using an autofocusing algorithm. This algorithm assigned zero phase to the raw holograms collected by the image sensor and propagated them to different sample to sensor heights using the free space angular spectrum approach [4]. The Tamura of the gradient (ToG) edge sparsity-based criterion was computed [38] for each hologram and used to calculate the corresponding refocusing distance.

For the lens-based diffraction-limited coherent imaging system (System B), the autofocusing algorithm required an additional background subtraction step. For undesired particles or dust associated with the objective lens or other parts of the optical microscope, the diffraction pattern that is formed is independent of the sample and its position. Using this information, a singular value decomposition (SVD)-based background subtraction was performed [39], after which the ToG-based autofocusing algorithm was successfully applied.

Multi-height phase recovery.

The iterative multi-height phase recovery technique [34] was applied to eliminate the holographic image artifacts (twin image and self-interference terms [4]) in both of the coherent imaging systems that were used in this work. To perform this, an initial zero-phase was assigned to the intensity/amplitude measurement at the 1st hologram height. Next, the iterative algorithm begins by propagating the complex field to each hologram height until the 8th height is reached, and then backpropagates the resulting fields until the 1st height is reached. While the phase was retained at each hologram height, the amplitude was updated by averaging current amplitude and the square root of the measured intensity at each height.

Registration between lower resolution and higher resolution (ground truth) images.

Image registration plays a key role in generating the training and testing image pairs for the network in both the pixel size-limited and diffraction-limited coherent imaging systems. A pixel-wise registration must be performed to ensure the success of the network in learning the transformation to perform super-resolution.

For both super resolution methods, the low-resolution input images were initially bicubically up-sampled. Following this, a correlation-based registration, which corrected any rotational misalignments or shifts between the images was performed. This registration process correlated the spatial patterns of the phase images and used the correlation to establish an affine transform matrix. This was in turn be applied to the high-resolution images to ensure proper matching of the corresponding fields-of-view between the low-resolution images and their corresponding ground truth labels. Finally, each image was cropped by 50 pixels to each side to accommodate for any relative shift that may have occurred.

For the diffraction-limited coherent imaging system (System B), an additional rough FOV matching step was required before the registration above. For this step, the higher resolution phase images was first stitched together, by calculating the overlap between neighboring images, and then fusing them together into a larger image. The corresponding lower resolution phase images are then matched to this larger image. This is done by creating a correlation score matrix between the large image and each smaller patch. Whichever portion of the matrix has the highest correlation score is used to determine which portion of the fused image is cropped out and is used as the input for the network.

GAN architecture and training process.

Once the high and low resolution image pairs were accurately registered, they were cropped into smaller image patches (128×128 pixels), which were used to train the network. The architectures of the generator (*G*) and the discriminator (*D*) that make up the GAN can be seen in Fig. 2.4.

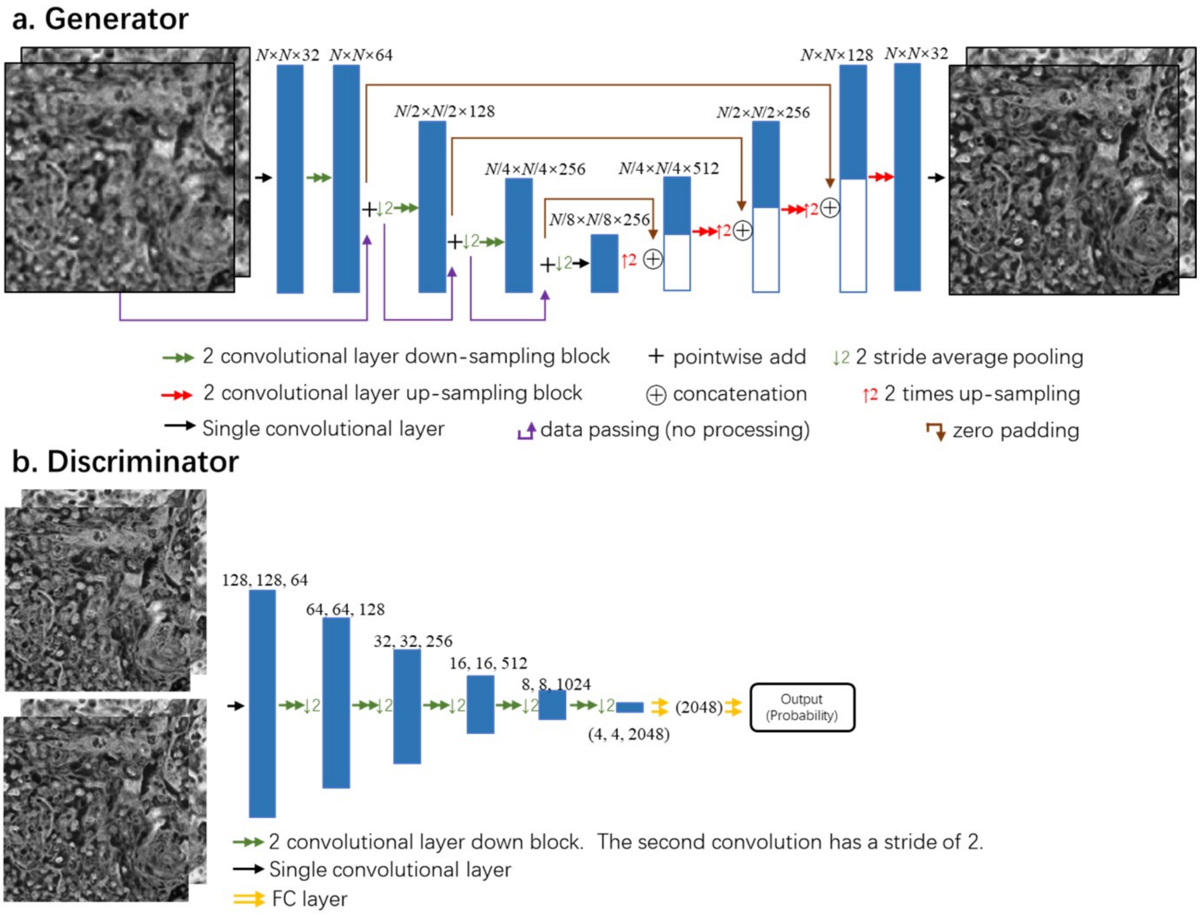


Figure 2.4 Diagram of the GAN structure. (a) Structure of the generator portion of the network. (b) Structure of the discriminator portion of the network.

For both the pixel-size limited and the diffraction-limited coherent imaging systems, the discriminator loss function is defined as:

$$l_{\text{discriminator}} = D\left(G\left(x_{\text{input}}\right)\right)^2 + \left(1 - D\left(z_{\text{label}}\right)\right)^2 \tag{4}$$

where $D(.)$ and $G(.)$ refer to the discriminator and generator network operators, respectively, x_{input} is the lower resolution input to the generator, and z_{label} is the higher resolution label image. For the lensfree holographic imaging system (System A), the generator loss function was defined by:

$$l_{\text{generator}} = L_1\left\{z_{\text{label}}, G\left(x_{\text{input}}\right)\right\} + \gamma \times TV\left\{G\left(x_{\text{input}}\right)\right\} + \alpha \times \left(1 - D\left(G\left(x_{\text{input}}\right)\right)\right)^2. \tag{5}$$

The $L_1\{z_{\text{label}}, G(x_{\text{input}})\}$ term is calculated using:

$$L_1\left\{z_{\text{label}}, G\left(x_{\text{input}}\right)\right\} = \mathrm{E}_{\text{n\_pixels}}\left(\mathrm{E}_{\text{n\_channels}}\left(\left|G\left(x_{\text{input}}\right) - z_{\text{label}}\right|\right)\right) \tag{6}$$

This finds the absolute difference between each pixel of the generator output image and its corresponding label. $\mathrm{E}_{\text{n\_pixels}}(.)$ and $\mathrm{E}_{\text{n\_channels}}(.)$ are the expectation values for the pixels within each image and the channels of each image, respectively. $TV\{G(x_{\text{input}})\}$ represents the total variation loss, which acts as a regularization term, applied to the generator output. Total variation (TV) is defined as:

$$TV = E_{\text{n\_channels}}\left(\sum\nolimits_{i,j}\left|G\left(x_{\text{input}}\right)_{i+1,j} - G\left(x_{\text{input}}\right)_{i,j}\right| + \left|G\left(x_{\text{input}}\right)_{i,j+1} - G\left(x_{\text{input}}\right)_{i,j}\right|\right) \tag{7}$$

where the i and j indices represent the location of the pixels within each channel of the image.

The last term in Equation (5) (i.e., $\alpha\times(1\text{-}D(\mathrm{G}(x_{\text{input}})))^2$) is a function of how well the output image of the generator network can be predicted by the discriminator network. α and γ are regularization parameters which were set to 0.00275 and 0.015 respectively. As a result of these parameters, the L_1 loss term, $L_1\{z_{\text{label}},G(x_{\text{input}})\}$, made up 60% of the overall loss, while the total variation term, $\gamma\times TV\{G(x_{\text{input}})\}$, was approximately 0.25% of the total loss. The discriminator loss term, $\alpha\times(1\text{-}D(G(x_{\text{input}})))^2$, made up the remainder of the overall generator loss. Once the networks were successfully trained, they reach a state of equilibrium where the discriminator network cannot successfully discriminate between the output and label images, and $D(G(x_{\text{input}}))$ converged to approximately 0.5.

The loss function for the lens-based coherent microscope images (System B) incorporated an additional structural similarity index (SSIM) [32] term in addition to the terms included for the lensfree on-chip imaging system, i.e.,:

$$\begin{aligned} l_{\text{generator}} = {} & L_1\left\{z_{\text{label}}, G\left(x_{\text{input}}\right)\right\} + \gamma\times TV\left\{G\left(x_{\text{input}}\right)\right\} + \alpha\times\left(1 - D\left(G\left(x_{\text{input}}\right)\right)\right)^2 \\ & + \beta\times\mathrm{SSIM}\left\{G\left(x_{\text{input}}\right), z_{\text{label}}\right\} \end{aligned} \tag{8}$$

Where β is a regularization parameter set to 0.01, and SSIM$\{x, z\}$ is defined as [32]:

$$\mathrm{SSIM}(x,z)=\frac{(2\mu_x\mu_z+c_1)(2\sigma_{x,z}+c_2)}{(\mu_x^2+\mu_z^2+c_1)(\sigma_x^2+\sigma_z^2+c_2)} \tag{9}$$

where μ_x, μ_z are the averages of x, z; σ_x^2,σ_z^2 are the variances of x, z, respectively; $\sigma_{x,z}$ is the covariance of x and z; and c_1, c_2 are dummy variables used to stabilize the division with a small denominator. The term $\beta\times\mathrm{SSIM}\{G(x_{\mathrm{input}}),z_{\mathrm{label}}\}$ was set to make up ~15% of the total generator loss, with the rest of the regularization weights reduced in value accordingly.

Our generator network used an adapted U-net architecture [40]. The network began with a convolutional layer that increased the number of channels to 32 and a leaky rectified linear (LReLU) unit, defined as:

$$\mathrm{LReLU}(x)=\begin{cases} x & \text{for } x>0 \\ 0.1x & \text{otherwise} \end{cases} \tag{10}$$

Following this layer, there was a down-sampling and an up-sampling section. Each section consists of three distinct layers, each made up of separate convolution blocks (see Fig. 2.4a). For the down-sampling section, these residual blocks consisted of two convolution layers with LReLU units acting upon them. At the output of the second convolution of each block the number of channels was doubled. The down-sampling blocks were connected by an average-pooling layer of stride two that down-samples the output of the previous block by a factor of two in both lateral dimensions (see Fig. 2.4a).

The up-sampling section of the network used a reverse structure to reduce the number of channels and return each channel to its original size. Similar to the down sampling section, each block contained two convolutional layers, each activated by a LReLU layer. At the input of each block, the previous output was up-sampled using a bilinear interpolation and concatenated with the output of the down-sampling path at the same level (see Fig. 2.4a). Between the two paths, convolutional layer was added to maintain the number of the feature maps from the output of the last residual block to the beginning of the down-sampling path

(Fig. 2.4a). Finally, a convolutional layer was used to reduce the number of output channels to match the size of the label.

The discriminator portion of the network was made up of a convolutional layer, followed by five discriminator blocks, an average pooling layer and two fully connected layers which reduced the output to a single value (see Fig. 2.4b). Both the label images and the output of the generator network were input into the initial convolutional layer discriminator network. This layer was used to increase the number of channels to 32 and was followed by five discriminator blocks all containing two convolutional layers activated upon by LReLU functions. The first convolution was used to maintain the size of the output, and the second doubled the number of channels while halving the size of the output in each lateral dimension. Next, the average pooling layer was used to find the mean of each channel, reducing the dimensionality to a vector of length 1024 for each patch. Each of these vectors were subsequently fed into two fully connected layers and LReLU activation layers in series. While the first fully connected layer did not change the dimensionality, the second reduced the output of each patch to a single number which was in turn input into a sigmoid function. The output of the sigmoid function represents the probability of the input being either real or fake and was used as part of the generator's loss function.

The filter size for each convolution was set to be 3×3. The trainable parameters are updated using an adaptive moment estimation (Adam) [41] optimizer with a learning rate 1×10^{-4} for the generator network and 1×10^{-5} for the discriminator network. The image data were augmented by randomly flipping 50% of the images, and randomly choosing a rotation angle (0, 90, 180, 270 degrees). For each iteration that the discriminator is updated, the generator network is updated four times, which helps the discriminator avoid overfitting to the target images. The convolutional layer weights are initialized using a truncated normal distribution while the network bias terms are initialized to zero. A batch size of 10 is used for the training, and a batch size of 25 is used for validation. The networks chosen for blind testing were those with the lowest validation loss. The number of training steps as well as the training time for each network are reported in Table 2.1, and the testing times are reported in Table 2.2.

Table 2.1 Training details for the deep neural networks.

Resolution limiting factor	Tissue type	Low resolution input type	Training dataset size (number of patches before augmentation)	Training time (s)	Number of iterations
Pixel size-limited (System A)	Pap smear	1×1 raw hologram	56250	9,078	17,000
	Lung	1×1 raw hologram	83700	17,052	28,000
	Lung	2×2 raw holograms	83700	9,363	15,000
	Lung	3×3 raw holograms	83700	30,480	52,500
Diffraction-limited (System B)	Pap smear	4×/0.13 NA objective lens	65475	46,411	100,000

All the networks were trained with a batch size of 10 using 128×128 pixel patches.

Table 2.2 Time for each network to output a 1940×1940 pixel image.

Resolution limiting factor	Tissue type	Low resolution input type	Testing Time (s/image)
Pixel size-limited (System A)	Pap smear	1×1 raw hologram	1.42
	Lung	1×1 raw hologram	1.37
	Lung	2×2 raw holograms	1.38
	Lung	3×3 raw holograms	1.38
Diffraction-limited (System B)	Pap smear	4×/0.13 NA objective lens	1.26

Each measurement is the average time, calculated using 150 test images.

Software implementation details.

The network was developed using a desktop computer running the Windows 10 operating system. The desktop uses an Nvidia GTX 1080 Ti GPU, a Core i7-7900K CPU running at 3.3 GHz, and 64 GB of RAM. The network was programmed using Python (version 3.6.0) with the TensorFlow library (version 1.7.0).

Sample preparation.

De-identified Pap smear slides were provided by the UCLA Department of Pathology (Institutional Review Board no. 11–003335) using ThinPrep® and SurePath™ preparation. De-identified Hematoxylin and Eosin (H&E) stained human lung tissue slides were acquired from the UCLA Translational Pathology Core Laboratory. We used existing and anonymous specimen, where no subject related information was linked or can be retrieved.

2.3 Results and discussion

Super-resolution of a pixel size-limited coherent imaging system. We first report the performance of the network for the pixel size-limited coherent imaging system using a Pap smear sample and a Masson's trichrome stained lung tissue section (connected tissue sample). For the Pap smear, two samples from different patients were used for training. For the lung tissue samples, three tissue sections from different patients were used for training. The networks were blindly tested on additional tissue sections from other patients. The FOV of each tissue image was ~20 mm^2 (corresponding to the sensor active area).

Fig. 2.5 illustrates the network's super-resolved output images along with pixel-size limited lower resolution input images and the higher resolution ground truth images of a Pap smear sample. The input images have a pixel pitch of 2.24 µm, and the label images have an effective pixel size of 0.37 µm (see the Methods section). For lung tissue sections, we also demonstrate our super-resolution results (Fig. 2.6) using three different deep networks, where the input images for each network has a different pixel size (i.e., 2.24 µm, 1.12 µm, and 0.7467 μm, corresponding to 1×1, 2×2 and 3×3 lateral shifts, respectively, as detailed in the Methods section). In comparison to the less densely connected Pap smear sample results, the network output misses some spatial details for lung tissue imaging when the input pixel size is at the coarsest level of 2.24 μm. These spatial features/details are recovered back by the other two networks that use smaller input pixels as shown in Fig. 2.6.

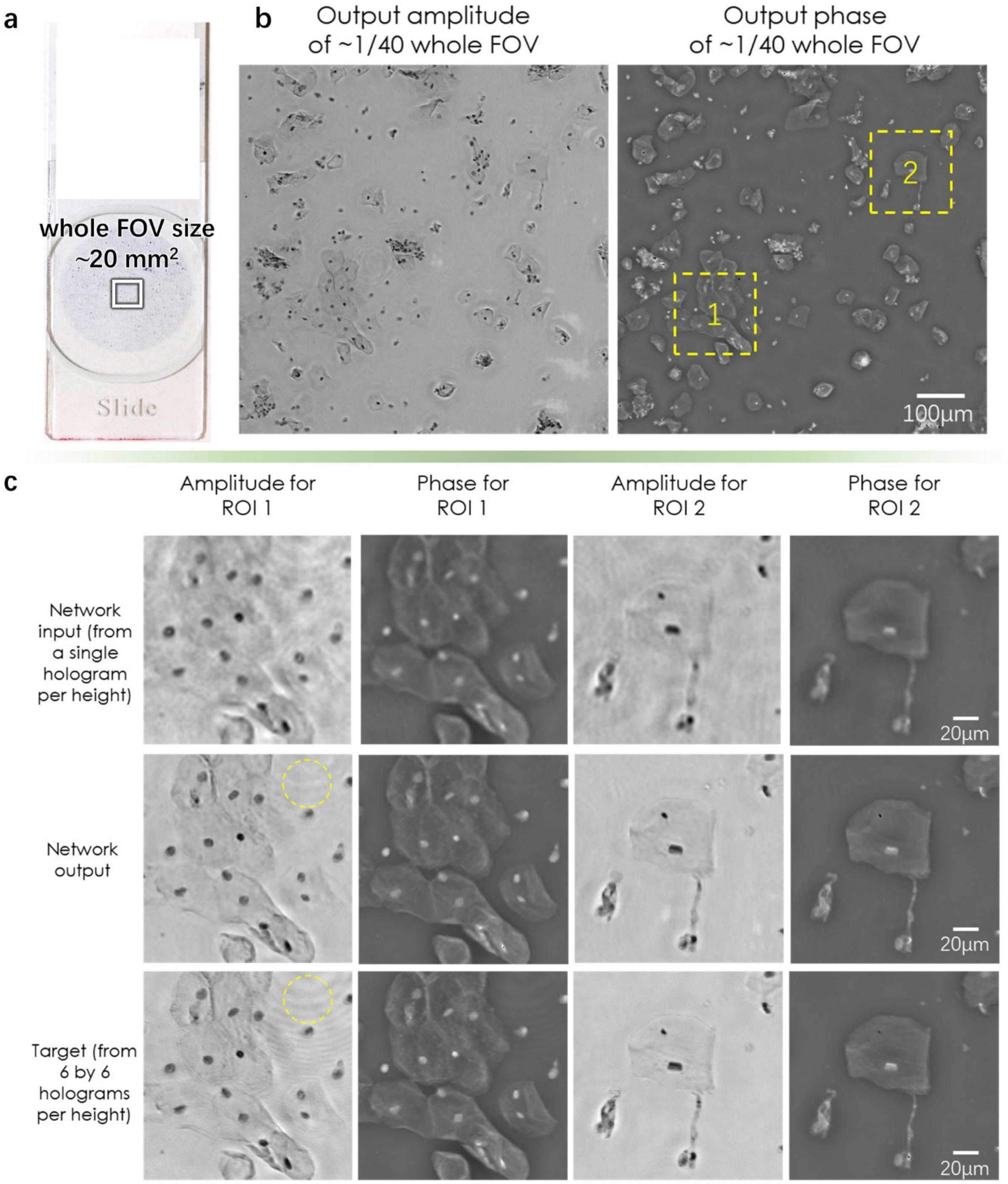


Figure 2.5 Visualized result for the pixel size-limited system. Deep learning-based pixel super-resolution imaging of a Pap smear slide under 550 nm illumination. (a) Whole FOV of the lensfree imaging system. (b) Amplitude and phase channels of the network output. (c) Further zoom-in of (b) for two regions of interest. The marked region in the first column demonstrates the network's ability to process the artifacts caused by out-of-focus particles within the sample.

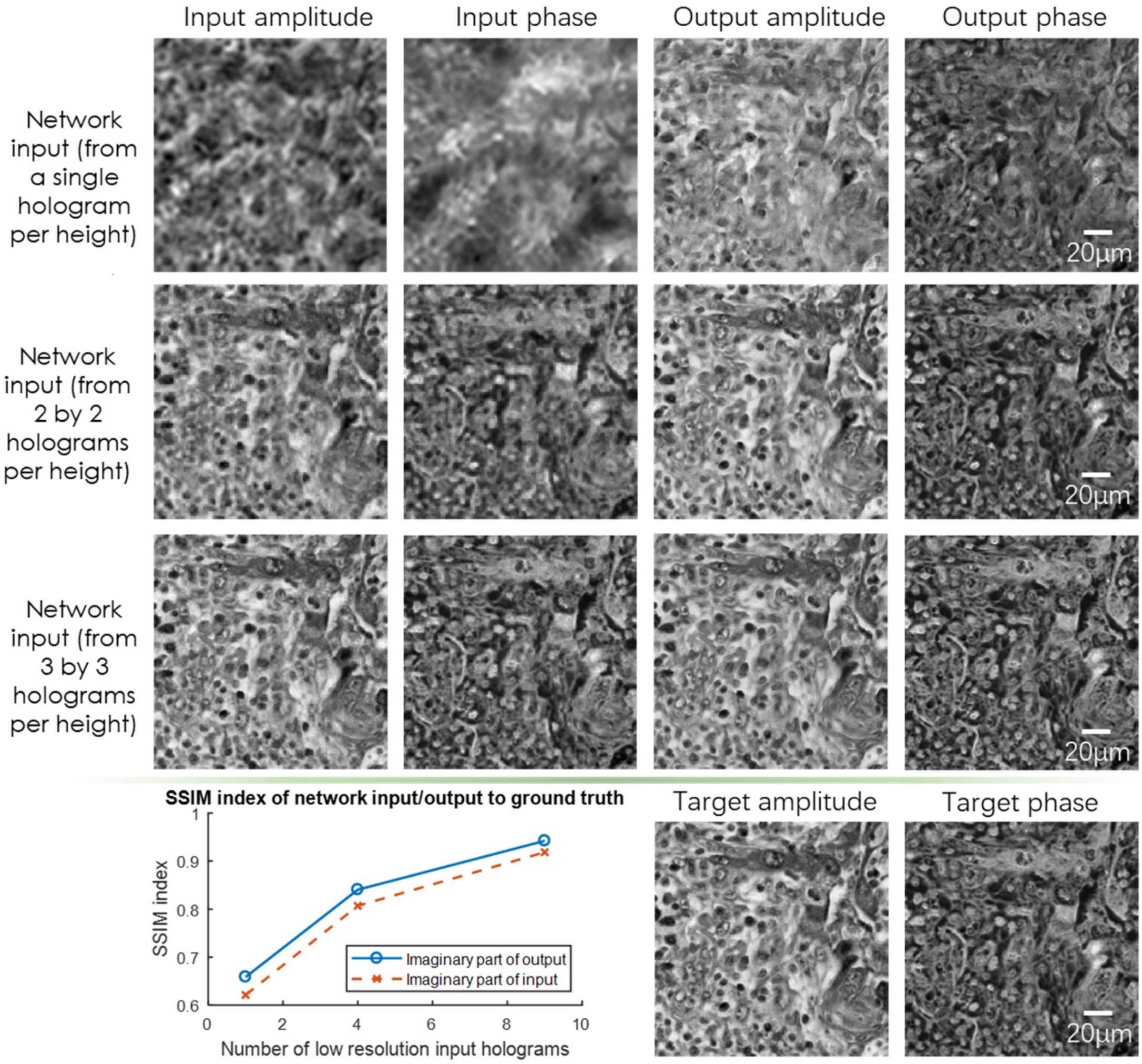


Figure 2.6 Visualized result for the pixel size-limited system. Comparison of the performances for the deep-learning-based pixel super-resolution methods using different input images. The sample is a Masson's trichrome stained lung tissue slide, imaged at an illumination wavelength of 550 nm. SSIM values are also shown for the network input and output images for each case. The ground truth (target) image for each SSIM value is acquired using 6×6 lensfree holograms per height.

We also report the SSIM values with respect to the reference label images in order to further evaluate the performance of our network output when applied to a pixel size-limited coherent imaging system. The average SSIM values for the entire image FOV (~20 mm^2) are listed in Table 2.3, where the input SSIM values were calculated between the bicubic interpolated lower resolution input images and the ground truth images. The results clearly demonstrate the improved structural similarity of the network output images.

Table 2.3 Average SSIM values for the lung and Pap smear samples for the deep neural network output (also see Figs. 2.5 and 2.6 for sample images in each category).

Resolution limiting factor	Tissue type	Low resolution input type	Input SSIM		Output SSIM	
			Imaginary	Real	Imaginary	Real
Pixel size-limited	Pap smear	1×1 raw hologram	0.9097	0.9135	0.9392	0.9442
	Lung	1×1 raw hologram	0.6213	0.5404	0.6587	0.7135
	Lung	2×2 raw holograms	0.8069	0.8205	0.8405	0.8438
	Lung	3×3 raw holograms	0.9185	0.9184	0.9422	0.9347

In addition to SSIM comparison, we also report the improved performance of our network output using spatial frequency analysis: Fig. 2.7 reports the 2-D spatial frequency spectra and the associated radially-averaged frequency intensity of the network input, network output and the ground truth images corresponding to our lensfree on-chip imaging system. The appearance of the higher spatial frequency components in the output of the network, approaching to the spatial frequencies of the ground truth image is another indication of our super-resolution performance.

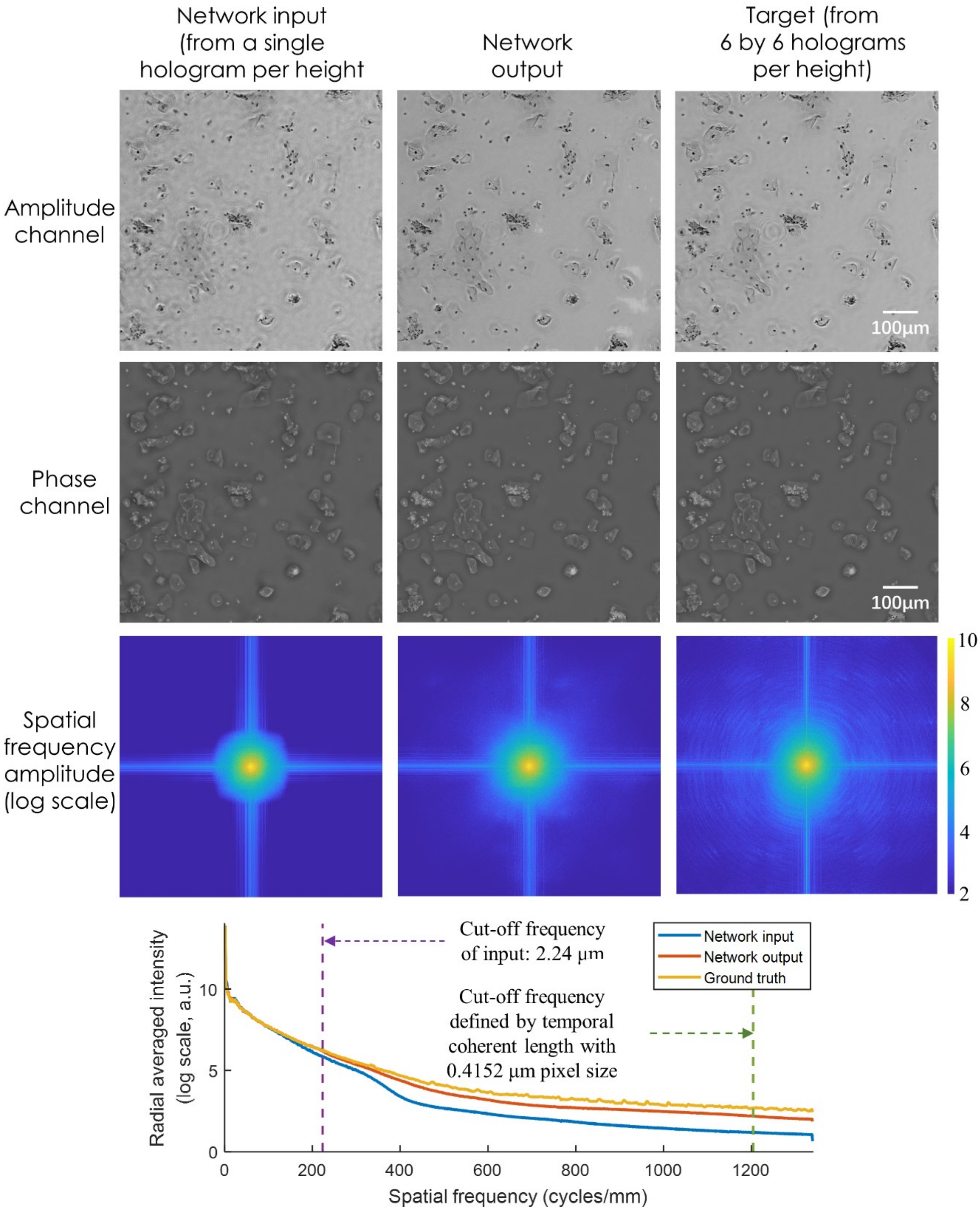


Figure 2.7 Spatial frequency analysis for the pixel size-limited system. Radially-averaged spatial frequency spectra of the network input, network output and target images, corresponding to a lensfree on-chip coherent imaging system.

Super-resolution of a diffraction-limited coherent imaging system. For the task of super-resolving a diffraction-limited coherent imaging system, we trained an identical network architecture (see the Methods section) with a Pap smear sample. As in the pixel super-resolution case reported earlier, two samples were obtained from two different patients, and the trained network was blindly tested on a third sample obtained from a third patient. The input images were obtained using a 4×/0.13 NA objective lens and the reference ground truth images were obtained by using a 10×/0.30 NA objective lens. Fig. 2.8 illustrates a visual comparison of the network input, output and label images, providing the same conclusions as in Fig. 2.5 and Fig. 2.6. Similar to the pixel size-limited coherent imaging system, we also analyzed the performance of our network using spatial frequency analysis which is reported in Fig. 2.9. The higher spatial frequencies of the network output image approach the spatial frequencies observed in the ground truth images, similar to the results of Fig. 2.7.

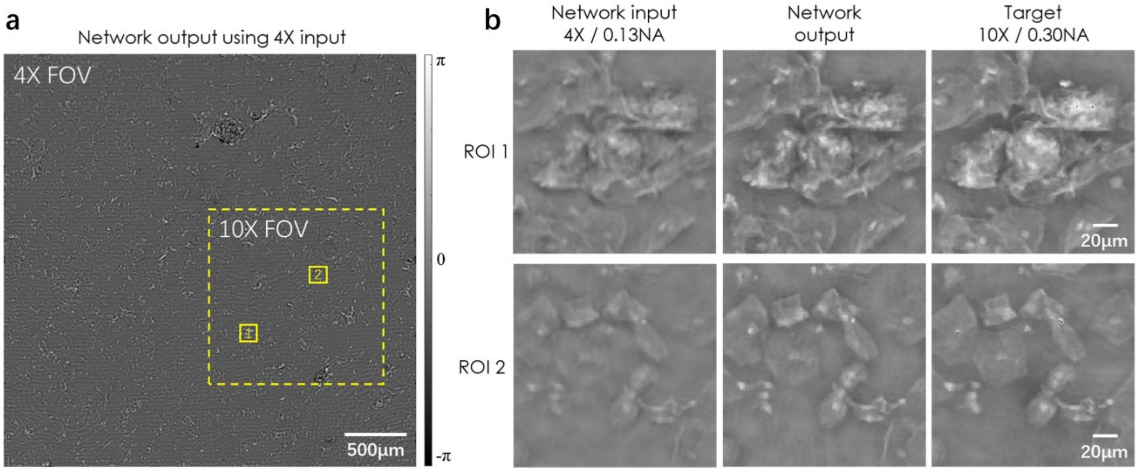


Figure 2.8 Visualized result for the diffraction-limited system. Deep learning-based super-resolution imaging of a Pap smear slide under 532 nm illumination using a lens-based holographic microscope. (a) Phase channel of the network output image. (b) Zoomed-in images of (a).

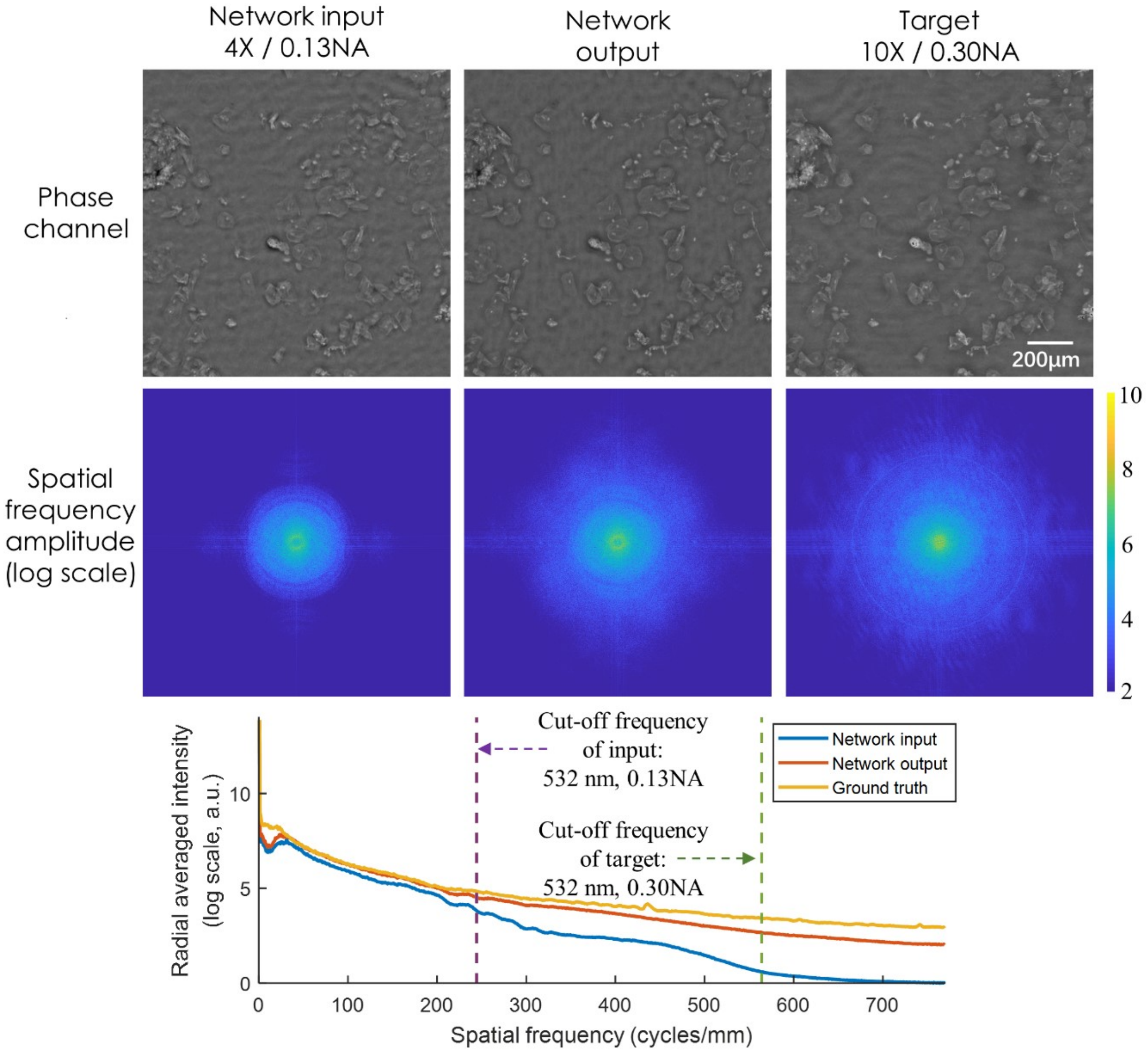


Figure 2.9 Spatial frequency analysis for the diffraction-limited system. Radially-averaged spatial frequency spectra of the network input, network output and target images, corresponding to a lens-based coherent imaging system.

On the other hand, the SSIM criterion did not reveal the same trend that we observed in the lensfree on-chip microscopy system reported earlier, and only showed a very small increase from e.g., 0.876 for the input image to 0.879 for the network output image. This is mostly due to increased coherence related artifacts and noise, compared to the lensfree on-chip imaging set-up, since the lens-based design has several optical components and surfaces within the optical beam path, making it susceptible to coherence induced background noise and related image artifacts, which partially dominate SSIM calculations.

2.4 Conclusions

We have presented a GAN-based framework for super-resolution of pixel size limited and diffraction limited coherent imaging systems. The framework was demonstrated on biologically connected thin tissue sections (lung and Pap smear samples) and the results were quantified using structural similarity index and spatial frequency spectra analysis. The proposed framework provides a highly optimized, non-iterative reconstruction engine that rapidly produces resolution enhancement, without additional parameter optimization.

The proposed approach is not restricted to a specific coherent imaging modality and is broadly applicable to various coherent image formation techniques. One of the techniques that will highly benefit from the proposed framework is off-axis holography. The proposed technique might be used to bridge the space-bandwidth-product gap between off-axis and in-line coherent imaging systems, while retaining the single-shot and high sensitivity advantages of off-axis image acquisition systems.

Chapter 3 . Deep learning-based color holography

3.1 Introduction

Histological staining of fixed, thin tissue sections mounted on glass slides is one of the fundamental steps required for the diagnoses of various medical conditions. Histological stains are used to highlight the constituent tissue parts by enhancing the colorimetric contrast of cells and subcellular components for microscopic inspection. Thus, an accurate color representation of the stained pathology slide is an important prerequisite to make reliable and consistent diagnoses [42–45]. Unlike bright-field microscopy, a common method used to obtain color information from a sample using a coherent imaging system requires the acquisition of at least three holograms at the red, green, and blue parts of the spectrum, thus forming the red–green–blue (RGB) color channels that are used to reconstruct composite color images. Such colorization methods used in coherent imaging systems suffer from color inaccuracies [46–48] and may be considered unacceptable for histopathology and diagnostic applications.

To achieve increased color accuracy using coherent imaging systems, a computational hyperspectral imaging approach can be used [49]. However, such systems typically require engineered illumination, such as a tunable laser to efficiently sample the visible band. Previous contributions have demonstrated successful reduction in the number of required sampling locations for the visible band to generate accurate color images. For example, Peercy *et al.* demonstrated a wavelength selection method using Gaussian quadrature or Riemann summation for reconstructing color images of a sample imaged in reflection mode holography [46], whereby it was suggested that a minimum of four wavelengths were required to generate accurate color images of natural objects. Later, Ito *et al.* demonstrated a Wiener estimation-based method to quantify the spectral reflectance distribution of the object at four fixed wavelengths that achieved an increased color accuracy for natural objects [50]. Recently, Zhang *et al.* presented an absorbance spectrum estimation method based on minimum mean-square-error estimation, specifically crafted to create accurate color images of pathology slides with in-line holography [48]. Because the color distribution within a stained histopathology slide is constrained by the colorimetric dye combination that is used, this method successfully reduced the required number of wavelengths to three, while it still preserved accurate color representation. However, owing to the distortions introduced by twin image artifacts and the limited resolution of unit magnification on-chip holography systems, multiheight phase recovery [5,7,33–36] and pixel super-resolution (PSR) techniques [51–61] were implemented to achieve acceptable image quality.

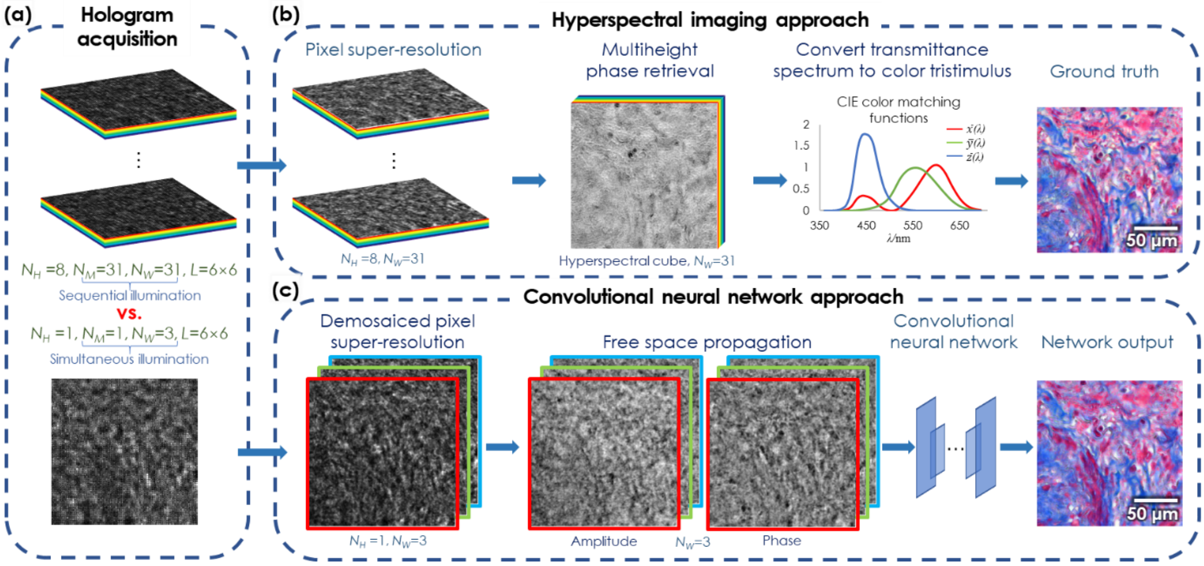


Figure 3.1 Comparison between the traditional hyperspectral imaging and the proposed neural network-based approaches for the reconstruction of accurate color images. N_H is the number of sample-to-sensor heights required for performing phase recovery, N_W is the number of illumination wavelengths, N_M is the number of measurements for each illumination condition (multiplexed or sequential), and L is the number of lateral positions used to perform pixel super resolution. **(a)**: Required number of raw holograms for the traditional hyperspectral imaging and the proposed neural network-based approaches. **(b)**: High fidelity color image reconstruction procedure for the hyperspectral imaging approach. **(c)**: High fidelity color image reconstruction procedure for the proposed neural network-based approach.

Herein, we present a deep learning-based accurate color holographic microscopy method **(Fig. 3.1)**. Neural networks and deep learning methods have been used in the literature to generate color images using a variety of imaging modalities. For example, grayscale photographs/images have been colorized using GANs [62,63]. As some other examples, holograms of objects (acquired at a single wavelength) have been reconstructed using deep neural networks with the color contrast of brightfield microscopy [64], and the images of unstained/label-free tissue samples have been transformed into brightfield equivalent color images of the same samples, demonstrating virtual staining of label-free tissue using holographic [65] or grayscale auto-fluorescence images of tissue sections [66]. In comparison to the traditional hyperspectral imaging approaches used in coherent imaging systems, the proposed deep neural-network-based color microscopy method of this work significantly simplifies the data acquisition procedures, the associated data processing and storage steps, and the imaging hardware. This technique requires only a *single* super-resolved hologram acquired under *wavelength-multiplexed* illumination. As such, the proposed approach achieves a similar performance to that of the state-of-the-art absorbance spectrum estimation method [48] that uses *four* super-resolved holograms collected at *four* sample-to-sensor distances with either sequential or multiplexed illumination wavelengths, thus representing more than four-fold enhancement in terms of data throughput.

We demonstrate the success of this framework using two types of pathology slides: lung tissue sections stained with Masson's trichrome and prostate tissue sections stained with Hematoxylin and Eosin (H&E). Using both the structural similarity index (SSIM) [67] and the color distance [68], high fidelity and color-accurate images are reconstructed and compared to the gold-standard images obtained using the hyperspectral imaging approach. The overall time performance of the proposed framework is also compared against a conventional 20× bright-field scanning microscope, thus demonstrating that the total image acquisition and processing times are of the same scale. We believe that the presented deep learning-based color imaging framework might be helpful to bring coherent microscopy techniques into use for histopathology applications.

Part of this chapter has been published in:

Liu, T., Wei, Z., Rivenson, Y., de Haan, K., Zhang, Y., Wu, Y., & Ozcan, A. (2019). Deep learning-based color holographic microscopy. Journal of biophotonics, 12(11), e201900107.

3.2 Materials and methods

3.2.1 Overview of the hyperspectral and deep neural network-based reconstruction approaches

We train a deep neural network to perform the image transformation from a complex field obtained from a single super-resolved hologram to the gold-standard image, which is obtained from $N_H \times N_M$ super-resolved holograms (N_H is the number of sample-to-sensor distances, and N_M is the number of measurements at one specific illumination condition). In this work, to generate the gold-standard images using the hyperspectral imaging approach, we used $N_H = 8$ and $N_M = 31$ *sequential* illumination wavelengths (ranging from 400 nm to 700 nm with 10 nm step size). The following subsections detail the procedures used to generate both the gold-standard images as well as the inputs to the deep network.

3.2.2 Hyperspectral imaging approach

The gold-standard, hyperspectral imaging approach reconstructs a high-fidelity color image by first performing resolution enhancement using a PSR algorithm (Section 3.2.2.1) Subsequently, the missing phase-related artifacts are eliminated using multiheight phase recovery (Section 3.2.2.3). Finally, high-fidelity color images are generated with tristimulus color projections (Section 3.2.2.4).

3.2.2.1 Holographic pixel super-resolution using sequential illumination

The resolution enhancement for the hyperspectral imaging approach was performed using a PSR algorithm [33]. This algorithm is capable of digitally synthesizing a high-resolution image (pixel size of approximately 0.37 μm) from a set of low-resolution images collected by an RGB image sensor (IMX 081, Sony, pixel size of 1.12 μm, with R, G_1, G_2, and B color channels). To acquire these images, the image sensor was programmed to raster through a 6×6 lateral grid using a 3D positioning stage (MAX606, Thorlabs, Inc.) with a subpixel spacing of ~0.37 μm (i.e., 1/3 of the pixel size). At each lateral position, one low-resolution hologram intensity was recorded. The displacement/shift of the sensor was accurately estimated using the algorithm introduced in [35]. A shift-and-add based algorithm was then used to synthesize the high-resolution image.

Because this hyperspectral imaging approach uses *sequential* illumination, the PSR algorithm uses only one color channel (R, G_1, or B) from the RGB image sensor at any given illumination wavelength. Based on the transmission spectral response curves of the Bayer RGB image sensor, the blue channel (B) was used for the illumination wavelengths in the range of 400–470 nm, the green channel (G_1) was used for the illumination wavelengths in the range of 480–580 nm, and the red channel (R) was used for the illumination wavelengths in the range of 590–700 nm.

3.2.2.2 Angular spectrum propagation

Free-space angular spectrum propagation [69] was used in the hyperspectral imaging approach to create the ground truth images. To digitally obtain the optical field $U(x,y;z)$ at a propagation distance z, the Fourier transform (FT) is first applied to the given $U(x,y;0)$ to obtain the angular spectrum distribution $A(f_x,f_y;0)$. The angular spectrum $A(f_x,f_y;z)$ of the optical field $U(x,y;z)$ can be calculated using:

$$A\left(f_x,f_y;z\right)=A\left(f_x,f_y;0\right)\cdot H\left(f_x,f_y;z\right) \tag{11}$$

where $H(f_x,f_y;z)$ is defined as,

$$H\left(f_x,f_y;z\right)=\begin{cases}0, & \left(\dfrac{\lambda f_x}{n}\right)^2+\left(\dfrac{\lambda f_y}{n}\right)^2>1\\ \exp\left[j2\pi\dfrac{n}{\lambda}z\sqrt{1-\left(\dfrac{\lambda f_x}{n}\right)^2-\left(\dfrac{\lambda f_y}{n}\right)^2}\right], & \text{otherwise}\end{cases} \tag{12}$$

where λ is the illumination wavelength, and n is the refractive index of the medium. Finally, an inverse Fourier transform is applied to $A(f_x,f_y;z)$ to get $U(x,y;z)$.

This angular spectrum propagation method first served as the building block of an autofocusing algorithm, which is used to estimate the sample to sensor distance for each acquired hologram [70,71]. After the accurate sample to the sensor distances were estimated, the hyperspectral imaging approach used the angular spectrum propagation as an additional building block for the iterative multiheight phase recovery, which will be detailed next.

3.2.2.3 Multiheight phase recovery

To eliminate the spatial image artifacts related to the missing phase, the hyperspectral imaging approach applied an iterative phase retrieval algorithm [34]. Holograms from eight sample-to-sensor distances were collected during the data acquisition step. The algorithm initially assigned a zero-phase to the intensity measurement of the object. Each iteration of the algorithm began by propagating the complex field from the first height to the eighth height, and by backpropagating it to the first height. The amplitude was updated at each height, while the phase was kept unchanged. The algorithm typically converged after 10–30 iterations. Finally, the complex field was backpropagated from any one of the measurement planes to the object plane to retrieve both the amplitude and the phase images.

3.2.2.4 Color tristimulus projection

Increased color accuracy was achieved by densely sampling the visible band at 31 different wavelengths in the range of 400 nm to 700 nm at a 10 nm step size. This spectral information was projected to a color tristimulus using the Commission Internationale de l’Éclairage (CIE) color matching function [47]. The color tristimulus in the XYZ color space can be calculated by,

$$\begin{aligned} X &= \int \bar{x}(\lambda)T(\lambda)E(\lambda)\mathrm{d}\lambda \\ Y &= \int \bar{y}(\lambda)T(\lambda)E(\lambda)\mathrm{d}\lambda \\ Z &= \int \bar{z}(\lambda)T(\lambda)E(\lambda)\mathrm{d}\lambda \end{aligned} \tag{13}$$

where λ is the wavelength, $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ are the CIE color matching functions, $T(\lambda)$ is the transmittance spectrum of the sample, and $E(\lambda)$ is the CIE standard illuminant D65 [47]. The XYZ values can be linearly transformed to the standard RGB values for display [47].

3.2.3 High-fidelity holographic color reconstruction via deep neural networks

The input complex fields for the proposed deep learning-based color reconstruction framework were generated in the following manner: Resolution enhancement and cross-talk correction through the demosaiced pixel super resolution algorithm (Section 3.2.3.1) followed by the initial estimation of the object via the angular spectrum propagation (Section 3.2.2.2).

3.2.3.1 Holographic demosaiced pixel super-resolution (DPSR) using multiplexed illumination

Similar to the hyperspectral imaging approach, the proposed network approach also used a shift-and-add-based algorithm in association with 6×6 low-resolution holograms to enhance the hologram resolution. We used three multiplexed wavelengths, i.e., simultaneously illuminated the sample with three distinct wavelengths. To correct the cross-talk error among different color channels in the RGB sensor we used the DPSR algorithm [61]. This cross-talk correction can be illustrated by the following equation:

$$\begin{bmatrix} U_{\mathrm{R}} \\ U_{\mathrm{G}} \\ U_{\mathrm{B}} \end{bmatrix} = \mathbf{W} \times \begin{bmatrix} U_{\mathrm{R\_ori}} \\ U_{\mathrm{G_1\_ori}} \\ U_{\mathrm{G_2\_ori}} \\ U_{\mathrm{B\_ori}} \end{bmatrix} \tag{14}$$

where $U_{\text{R-ori}}$, $U_{\text{G}_1\text{-ori}}$, $U_{\text{G}_2\text{-ori}}$, and $U_{\text{B-ori}}$, represent the original interference patterns collected by the image sensor, **W** is a 3×4 cross-talk matrix obtained by experimental calibration of a given RGB sensor chip, and U_{R}, U_{G}, and U_{B}, are the demultiplexed (R, G, B) interference patterns. In this work, the three illumination wavelengths were chosen to be at 450 nm, 540 nm, and 590 nm. As suggested in [48], using these wavelengths, a better color accuracy can be achieved with specific tissue-stain types (i.e., prostate stained with H&E and lung stained with Masson's trichrome, which were used in this work).

3.2.3.2 Deep neural network input formation

Following the demosaiced pixel-super-resolution algorithm, the three intensity holograms are numerically backpropagated to the object plane, as discussed in Subsection 3.2.2.2. Following this back-propagation step, each one of the three color hologram channels will produce a complex wave, represented as real and imaginary data channels. This results in a six-channel tensor that is used as input to the deep network, as shown in **Fig. 3.1**. Unlike the ground truth, in this case, no phase retrieval is performed because only a single measurement is available.

3.2.3.3 Deep neural network architecture

A generative adversarial network (GAN [72]) was implemented to learn the color correction and eliminate the missing phase-related artifacts. This GAN framework has recently found applications in super-resolution microscopic imaging [30,73,74] and histopathology [64,66], and it consists of a discriminator network (*D*) and a generator network (*G*). The *D* network was used to distinguish between a three-channel RGB ground truth image (z) obtained from hyperspectral imaging and the output image from *G*. Accordingly, *G* was used to learn the transformation from a six-channel holographic image (x), i.e., three color channels with real and imaginary components, into the corresponding RGB ground truth image.

Our discriminator and generator losses are defined as,

$$l_{\text{discriminator}} = D\left(G\left(x\right)\right)^2 + \left(1 - D\left(z\right)\right)^2 \tag{5}$$

$$l_{\text{generator}} = L_2\left\{z, G\left(x\right)\right\} + \lambda \times TV\left\{G\left(x\right)\right\} + \alpha \times \left(1 - D\left(G\left(x\right)\right)\right)^2 \tag{6}$$

where,

$$L_2\left\{z, G\left(x\right)\right\} = \frac{1}{N_{\text{channels}} \times M \times N} \sum_{n=1}^{N_{\text{channels}}} \sum_{i,j=1}^{M,N} (x_{i,j,n} - z_{i,j,n})^2 \tag{7}$$

where N_{channels} is the number of channels in the images (e.g., $N_{\text{channels}} = 3$ for an RGB image), M and N are the number of pixels for each side of the images, i and j are the pixel indices, and n denotes the channel indices. TV represents the total variation regularizer that applies to the generator output, and is defined as,

$$TV(x) = \frac{1}{N_{\text{channels}}} \sum_{n=1}^{N_{\text{channels}}} \sum_{i,j=1}^{M,N} \left(\left| x_{i+1,j,n} - x_{i,j,n} \right| + \left| x_{i,j+1,n} - x_{i,j,n} \right| \right) \tag{8}$$

The regularization parameters (λ, α) were set to 0.0025 and 0.002 so that the total variation loss ($\lambda \times TV\{G(x_{\text{input}})\}$) is ~2% of L_2, and the discriminator loss ($\alpha \times (1\text{-}D(G(x_{\text{input}})))^2$) is ~15% of $l_{\text{generator}}$. Ideally, both $D(z_{\text{label}})$ and $D(G(x_{\text{input}}))$ converge to 0.5 at the end of the training phase. The L2-loss was empirically found to better handle distortions, which resulted due to the free-space back propagation of the single RGB DPSR hologram.

The generator network architecture (**Fig. 3.2**) was an adapted form of the U-net [75]. Additionally, the discriminator network (see **Fig. 3.3**) used a simple classifier that consisted of a series of convolutional layers which slowly reduced the dimensionality, while they increased the number of channels, followed by two fully connected layers to output the classification. While in this manuscript we adapted the U-net structure for our deep network, other structures can also be considered for elimination of missing phase artifacts [76] and for performing color correction on the reconstructed images. The convolution filter size was set to 3×3, and each convolutional layer except the last was followed by a leaky-ReLu activation function, defined as:

$$leaky\text{-}ReLu(x)=\begin{cases} x & \text{for } x>0 \\ 0.1x & \text{otherwise} \end{cases} \quad (9)$$

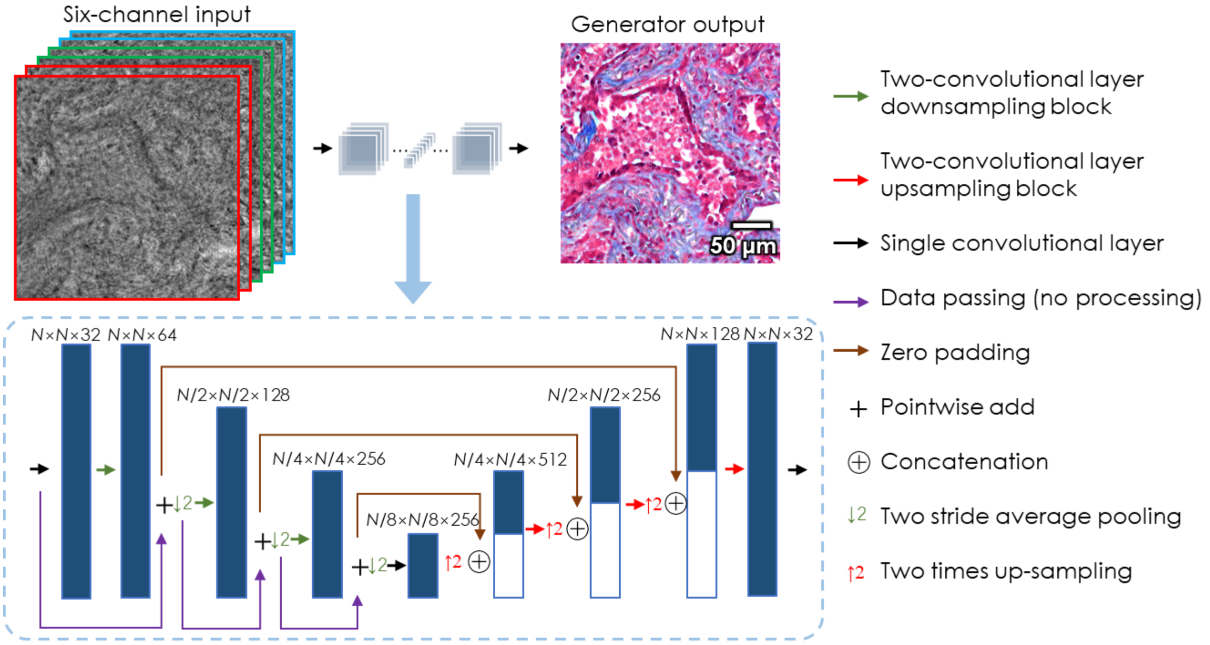


Figure 3.2 Schematic of the generator part of the network. The six-channel input consists of the real and imaginary channels of the three free-space propagated holograms at three illumination wavelengths (450 nm, 540 nm, and 590 nm). Each down block consists of two convolutional layers that double the number of system channels when used together. The down blocks are opposite, and consist of two convolutional layers with half the number of system channels when used together.

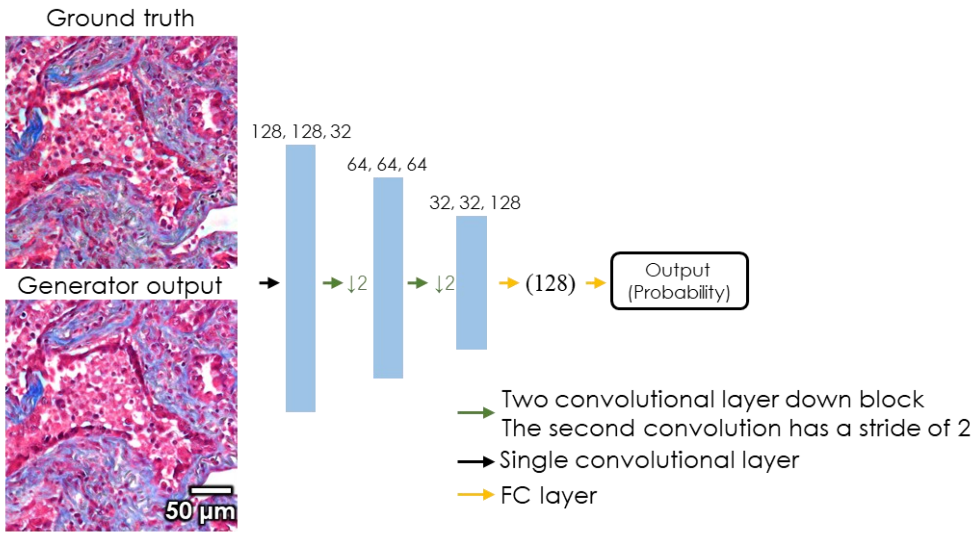


Figure 3.3 Diagram of the discriminator part of the network. Each down block of the convolutional layer consists of two convolutional layers.

3.2.3.4 Deep neural network training process

In the network training process, we used the images generated by the hyperspectral approach as our network labels, and took the demosaiced super-resolved holograms that were back-propagated to the sample plane as the network inputs. Both the generator and the discriminator networks were trained with non-overlapping patches, each with a size of 128×128 pixels. The weights in the convolutional layers and fully connected layers, were initialized using the Xavier initialization [77] while the biases were initialized to zero. All parameters were updated using an adaptive moment estimation (Adam) optimizer [78] with a learning rate of 1×10^{-4} for the generator network and a corresponding rate of 5×10^{-5} for the discriminator network. The training, validation, and testing of the network were performed on a PC with a four-core 3.60 GHz CPU, 16 GB of RAM, and an Nvidia GeForce GTX 1080Ti GPU. The lung tissue network was trained for 38.9 epochs over 5.58 hours, while the prostate tissue network was trained for 25.6 epochs over 2.29 hours. This training process only needs to be performed once for a specific type of tissue-stain combination and can improve in speed by using cloud computing.

3.2.4. Bright-field imaging

For comparison of the imaging throughput, bright-field microscopy images were obtained. An Olympus IX83 microscope equipped with a motorized stage and a set of super panchromatic objectives (Olympus UPLSAPO 20×/0.75 numerical aperture (NA), working distance (WD) 0.65) were used. The microscope was controlled by the MetaMorph advanced digital imaging software (Version 7.10.1.161, MetaMorph®) with the autofocusing algorithm set to search in a range of 5 μm in the z–direction with 1 μm accuracy. Two-pixel binning was enabled and a 10% overlap between the scanned patches was used. Stitching was done using the ImageJ Grid/Collection stitching plugin [79], which calculates the exact overlap between the images, and linearly blends the overlapping section, which allows the image to have a smooth transition and reduces stitching related artifacts.

3.2.5. Quantification metrics

Quantification metrics were chosen and used to evaluate the performance of the network: the SSIM [67] was used to compare the similarity of the tissue structural information between the output and the target images; ΔE*94 [68] was used to compare the color distance of the two images.

SSIM values ranged from zero to one, whereby the value of unity indicated that the two images were the same, i.e.,

$$\mathrm{SSIM}(U,V)=\frac{(2\mu_U\mu_V+C_1)(2\sigma_{U,V}+C_2)}{(\mu_V^2+\mu_V^2+C_1)(\sigma_U^2+\sigma_V^2+C_2)} \tag{10}$$

where U and V represent one vectorized test image and one vectorized reference image, respectively, μ_U and μ_V are the means of U and V, respectively, σ_U^2, σ_V^2 are the variances of U and V, respectively, $\sigma_{U,V}$ is the covariance of U and V, and constants C_1 and C_2 are included to stabilize the division when the denominator is close to zero.

The second metric that we used, ΔE*94 [68], outputs a number between zero and 100. A value of zero indicates that the compared pixels share the exact same color, while a value of 100 indicates that the two images have the opposite color (mixing two opposite colors cancel each other out and produce a grayscale color). This method calculates the color distance in a pixel-wise fashion, and the final result is calculated by averaging the values of ΔE*94 in every pixel of the output image.

3.2.6. Sample preparation

De-identified H&E stained human prostate tissue slides and Masson's trichrome stained human lung tissue slides were acquired from the UCLA Translational Pathology Core Laboratory. Existing and anonymous specimens were used. No subject related information was linked or can be retrieved.

3.3 Results and discussion

3.3.1. Qualitative assessment

We evaluated our network's performance using two different tissue-stain combinations: prostate tissue sections stained with H&E, and lung tissue sections stained with Masson's trichrome. For both types of samples, the networks were trained on three tissue sections from different patients and were blindly tested on another tissue section from a fourth patient. The field-of-view (FOV) of each tissue section that was used for training and testing was ~20 mm^2. The results for lung and prostate samples are respectively summarized in **Fig. 3.4** and **3.5**. These indicate our approach's capability of reconstructing a high-fidelity and color-accurate image from a single nonphase-retrieved and wavelength-multiplexed hologram (as detailed in the Methods section). Using the trained model, we were able to reconstruct the sample image over the entire sensor's FOV (i.e., ~20 mm^2), as demonstrated in **Fig. 3.6**.

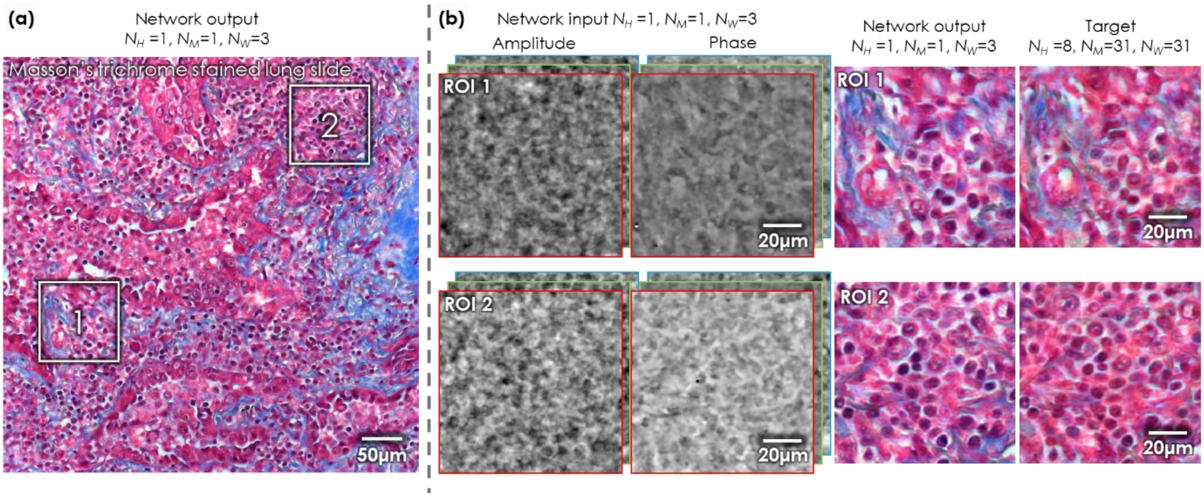


Figure 3.4 Deep learning-based accurate color imaging of a lung tissue slide stained with Masson's trichrome for a multiplexed illumination at 450 nm, 540 nm, and 590 nm, using a lens-free holographic on-chip microscope. (a): Large field of view of the network output image. (b): Zoomed-in comparison of the network input, the network output, and the ground truth target at region of interest (ROI) 1 and 2.

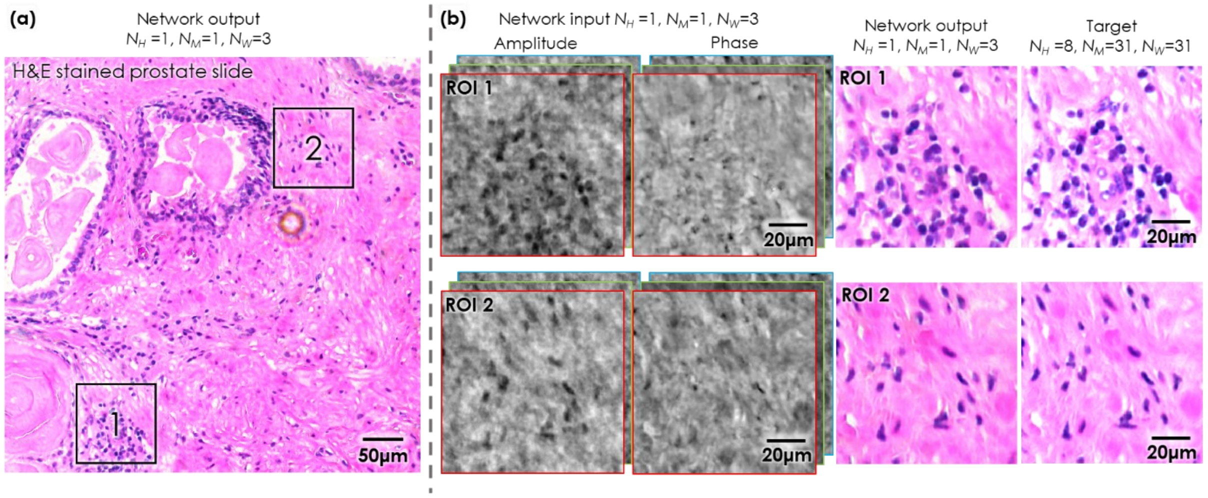


Figure 3.5 Deep learning-based accurate color imaging of a prostate tissue slide stained with H&E for a multiplexed illumination at 450 nm, 540 nm, and 590 nm, using a lens-free holographic on-chip microscope. (a): Large field of view of the network output image. **(b)** Zoomed-in comparison of the network input, the network output, and the ground truth target at ROI 1 and 2.

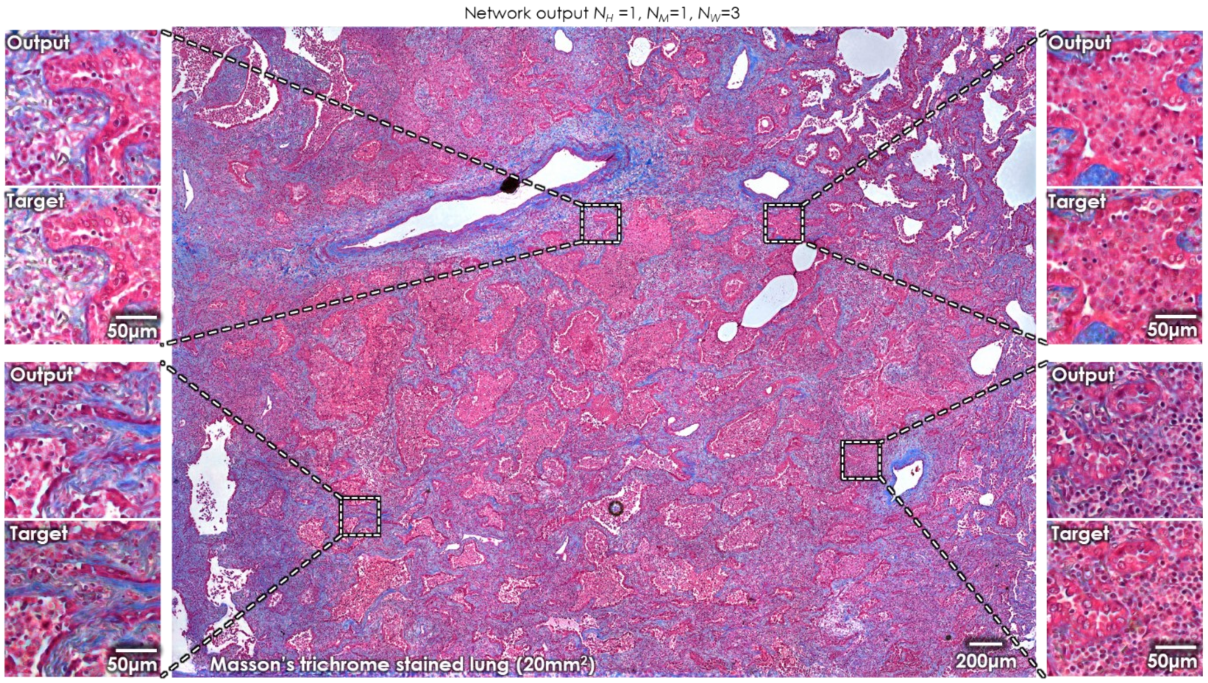


Figure 3.6 Stitched image of the deep neural network output for a lung tissue section stained with H&E, which corresponds to the sensor's field-of-view.

To further demonstrate the qualitative performance of the network, we compare in **Fig. 3.7** and **3.8** the reconstruction results of the deep network to the images created by the absorbance spectrum estimation method [48] in terms of the required number of measurements. For this comparison, we implemented the spectrum estimation approach for the multiheight phase recovery method and reconstructed the color images from a reduced number of wavelengths via both sequential (N_H=8, N_M=3) and multiplexed (N_H=8, N_M=1) illuminations at the same wavelengths (i.e., 450 nm, 540 nm, and 590 nm). Qualitatively, the network results are comparable to the multiheight results obtained with more than four sample-to-sensor distances for both the sequential and multiplexed illumination cases. This will be also confirmed by the quantitative analysis described below.

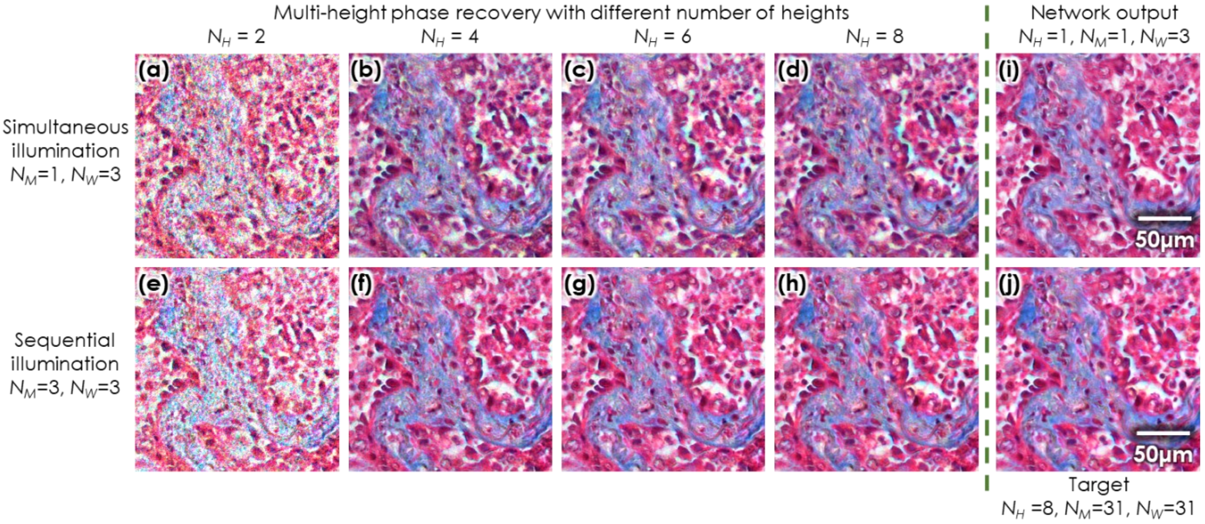


Figure 3.7 Visual comparison between the deep neural network-based approach and the multiheight phase recovery with spectral estimation approach for a lung tissue sample stained with Masson's trichrome. (a-h): Reconstruction results of spectral estimation approach using different number of heights and different illumination conditions. (i): Network output. (j): Ground truth target obtained using the hyperspectral imaging approach

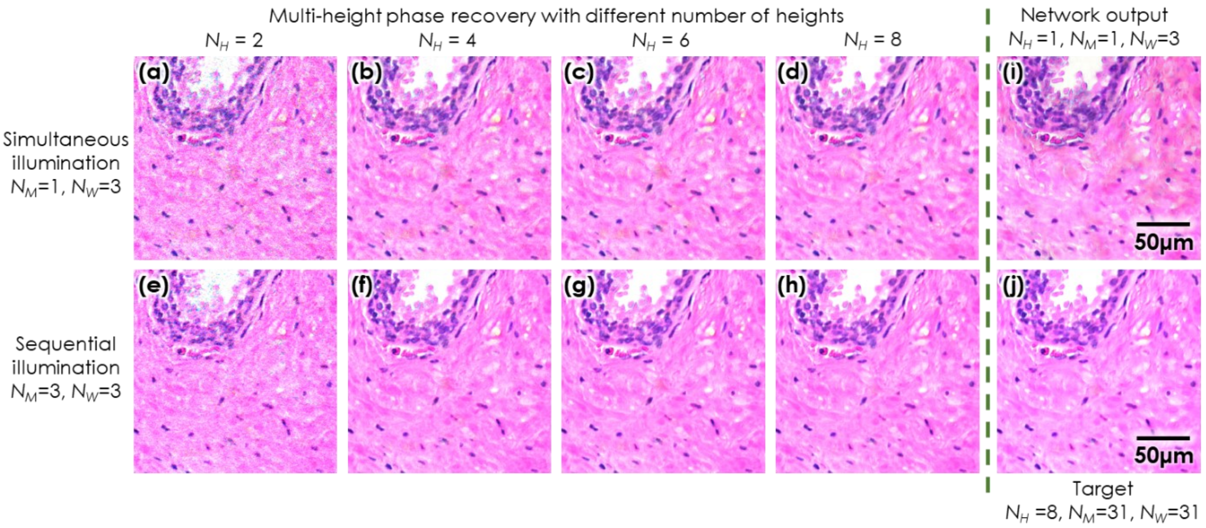


Figure 3.8 Visual comparison between the deep neural network-based approach and the multiheight phase recovery with the spectral estimation approach for a prostate tissue sample stained with H&E. (a-h): Reconstruction results of spectral estimation approach using different number of heights and different illumination conditions. (i): Network output. (j): Ground truth target obtained using the hyperspectral imaging approach

3.3.2. Quantitative performance assessment

The quantitative performance of the network was evaluated based on the calculation of the SSIM [67] and color difference (ΔE*94 [68]) between the network's output and the gold-standard image produced by the hyperspectral imaging approach. As listed in Table 3.1 and visually shown in **Fig. 3.5 and 3.6**, the performances of the spectrum estimation methods decrease (i.e., SSIM decreases and ΔE*94 increases) as the number of holograms at different sample-to-sensor distances decreases, or when the illumination is changed to be multiplexed. This quantitative comparison demonstrates that the network's performance using a single super-resolved hologram is comparable to the results obtained by state-of-the-art algorithms where ≥4 times as many raw holographic measurements are used.

Table 3.1 Comparison of SSIM) and ΔE*94 performances between the deep neural network approach and various other methods using two, four, six, and eight sample-to-sensor heights and three sequential/multiplexed wavelength illumination conditions for two tissue samples (the network-based approach and other methods with comparable performance are highlighted with bold font).

Tissue-stain type	**Method**	**Illumination condition (at 450 nm, 540 nm, and 590 nm)**	**Total required measurements ($N_H \times N_M \times L$)**	**Average SSIM**	**ΔE*94**
Masson's trichrome stained lung slide (~20 mm^2 FOV)	Deep neural network	Simultaneous	1×1×36	**0.8396**	6.9044
	Two-height reconstruction	Simultaneous	2×1×36	0.5535	10.7507
		Sequential	2×3×36	0.6011	9.4786
	Four-height reconstruction	Simultaneous	4×1×36	**0.8344**	5.1674
		Sequential	4×3×36	**0.8769**	3.8709
	Six-height reconstruction	Simultaneous	6×1×36	**0.878**	4.4219
		Sequential	6×3×36	0.9136	3.1928
	Eight-height reconstruction	Simultaneous	8×1×36	0.9068	3.6779
		Sequential	8×3×36	0.9538	2.1849
Hematoxylin and Eosin stained prostate slide (~20 mm^2 FOV)	Deep neural network	Simultaneous	1×1×36	**0.9249**	4.5228
	Two-height reconstruction	Simultaneous	2×1×36	0.7716	7.5085
		Sequential	2×3×36	0.848	5.5316
	Four-height reconstruction	Simultaneous	4×1×36	**0.8984**	4.3878
		Sequential	4×3×36	**0.9335**	3.3399
	Six-height reconstruction	Simultaneous	6×1×36	**0.9225**	3.8911
		Sequential	6×3×36	0.9516	2.9622
	Eight-height reconstruction	Simultaneous	8×1×36	0.9411	3.5102
		Sequential	8×3×36	0.9689	2.4148

3.3.3. Throughput evaluation

Table 3.2 lists the measured reconstruction times for the entire FOV (~20 mm^2) using different methods. For the deep neural network approach, the total reconstruction time includes the acquisition of 36 holograms (at 6×6 lateral positions in multiplexed illumination), the execution of DPSR, angular spectrum propagation, network inference, and image stitching. For the hyperspectral imaging approach, the total reconstruction time includes the collection of 8928 holograms (at 6×6 lateral positions, eight sample-to-sensor distances, and 31 wavelengths), PSR, multiheight phase retrieval, color tristimulus projection, and image stitching. For the conventional bright-field microscope (equipped with an automatic scanning stage), the total time includes the scanning of the bright-field images using a 20×/0.75 NA microscope with autofocusing performed at each scanning position and image stitching. In addition, the timing of the multiheight phase recovery method with the use of four sample-to-sensor distances was also shown, and had the closest performance to the deep learning-based neural network approach. All the coherent imaging related algorithms were accelerated with an Nvidia GTX 1080Ti GPU and CUDA C++ programming.

Table 3.2 Time performance evaluation of the deep neural network approach for reconstructing accurate color images compared to traditional hyperspectral imaging approach and standard brightfield microscopic sample scanning (where N/A stands for "not applicable").

Testing area	**Method**	**Data acquisition time**	**Processing time**					**Total time**	**Storage space (raw data)**
			Auto-Focusing	**Super resolution**	**Phase recovery or FSP**	**Inference or color transformation**	**Stitching**		
Sensor's entire FOV ~20 mm^2	Deep neural network	~2 min	~ 20 s	~2 min	~ 3 s	~ 1.5 min	~1 min	~7 min	1.09 GB
	Four-height simultaneous	~ 8 min	~ 80 s	~ 9 min	~ 5 min	~36 min	~1 min	~60 min	4.36 GB
	Four-height sequential	~ 25 min	~ 80 s	~ 9 min	~ 5 min	~36 min	~1 min	~77 min	13.08 GB
	Hyperspectral imaging	~ 8 h	~ 27 min	~ 3 h	~ 85 min	~15 min	~1 min	~13 h	270.32 GB
	Conventional microscope (20×/0.75 NA)	~6 min	N/A	N/A	N/A	N/A	~1 min	~7 min	577.13 MB

The network-based method took ~7 min to acquire and reconstruct a 20 mm^2 tissue area, which was approximately equal to the time it would take to image the same region using the 20× objective with our standard, general-purpose, bright-field scanning microscope. This is significantly shorter than the ~60 min required when using the spectral estimation approach (with four heights and simultaneous illumination). The deep learning approach also increases the data efficiency. The raw super-resolved hologram data size was reduced from 4.36 GB to 1.09 GB, which is more comparable to the data size of bright-field scanning microscopy images, which in total used 577.13 MB.

3.4 Conclusions

We presented a deep learning-based color holographic imaging system and demonstrated its performance using histologically stained pathology slides. This framework significantly simplified the data acquisition procedure, reduced the data storage requirement, shortened the processing time, and enhanced the color accuracy of the holographically reconstructed images. Here, we demonstrated the effectiveness of the presented approach using deep neural networks trained with specific sample-stain combinations that were holographically imaged. Therefore, the general recommendation is to use a separate trained network for a different type of stain; while this is a limitation, it does not affect the potential benefits of the technique for the targeted applications in e.g., pathology, where the type of the sample is fixed and known a priori. It is also important to note that other technologies, such as slide-scanner microscopes used in pathology can readily scan tissue slides at much faster rates, although they are rather expensive for use in resource limited settings. Therefore, further improvements to our lensless holographic imaging hardware, such as for example, the use of illumination arrays to perform pixel super resolution [54] would be needed to improve the overall reconstruction time of our results.

Chapter 4. Holographic polarization microscopy using deep learning

4.1 Introduction

Polarized light microscopy is widely used as a diagnostic tool in pathology, as it introduces distinctive contrast to birefringent specimen [80]. A number of diseases, such as squamous cell carcinoma [81], primary cutaneous amyloidosis [82], cerebral amyloid angiopathy [83], and senile cardiovascular amyloid [84] can be diagnosed using various polarization imaging techniques. Since 1961, compensated polarized light microscopy (CPLM) has been the gold standard imaging technique to identify monosodium urate (MSU) [85] crystals in synovial fluid samples [86], and is used to diagnose gout and pseudogout [87]. CPLM operates by allowing linearly polarized white light illumination to pass through a full-waveplate designed for green light (commonly between 530 nm to 560 nm), which in combination with a linear polarizer/analyzer, generates a magenta background. The presence of a birefringent specimen within the light path changes the polarization state of the green light, which shifts the spectrum after the analyzer and results in the final image becoming yellow or blue.

While CPLM images are treated as the gold standard for MSU crystal detection, the effort is labor intensive as microscopes have limited fields-of-view (FOV) and therefore, mechanical scanning is required to inspect the whole sample area. In addition to having a limited FOV, the CPLM technique suffers from limitations such as requiring manual alignment of the polarizer in relation to the analyzer, limited focal depth with higher magnification objectives and limited sensitivity when being used to detect small crystals or crystals with weak birefringence. As a result, CPLM analysis is sensitive to both the concentration of the crystals in synovial fluid [88] and experience of the diagnostician/technician [89]. Finally, clinical CPLM reporting is limited to *qualitative* results (e.g., presence or absence of crystals on the slide and whether crystals are intra- or extra-cellular).

There are a number of alternative polarization microscopy methods which have been developed to produce quantitative images of birefringent specimen. These methods all use the same principles of operation: they collect images from two or more light paths with either the polarizer or the analyzer oriented differently in order to infer the Stokes/Jones parameters that define the birefringent specimen [90–94]. Among these methods, Jones phase microscopy [93] and polarization holographic microscopy, (PHM) [94] are two examples, where both systems take advantage of the amplitude and phase information of the reconstructed interferogram, and measure spatially resolved Jones matrix components of anisotropic samples using four polarization states. However, these types of systems in general require the use of relatively sophisticated and costly optical components to maintain a linear mapping between the measurements and the inferred Jones parameters.

Various methods have recently demonstrated that the retardance and orientation information channels, especially for biological samples, provide some of the most useful spatial features corresponding to birefringent specimen [95–97]. Due to the reduction of the number of unknown parameters, these methods usually have simpler optical designs compared to e.g., PHM. Among these, a technique known as single-shot computational polarized light microscopy (SCPLM) [97] that uses a pixel-wise polarized image sensor with four polarization directions has been demonstrated to simplify the optical system required to image birefringent samples. Using this method, the retardance and orientation of the sample are explicitly solved, providing quantitative contrast for birefringent specimen. While these methods are quite effective, both PHM and SCPLM are lens-based imaging systems. Therefore, they suffer from the small field of view of objective lenses and a relatively high system cost. One method which can get past these limitations is wide-field lensfree differential holographic polarized microscopy [98,99]. By taking advantage of the simple optical design and unit magnification of lensfree on-chip holographic systems, this method can achieve a FOV of >20-30 mm^2 [100–102]. Furthermore, the technique is cost-effective, compact and suitable for resource-limited settings [103,104]. However, in order for this lensfree holographic imaging method to be used for imaging of birefringent objects, two sets of raw holograms must be taken with illuminations in two different polarization states, which requires precise image alignment, especially for the detection of small birefringent objects within the sample. Furthermore, this method does not resolve the retardance or orientation of the sample, leading to lower contrast compared to the SCPLM method.

In parallel to these advances in computational polarization microscopy, deep learning has emerged as a highly effective technique for solving inverse problems in microscopy [105,106]. It has been applied to traditional inverse problems such as holographic image reconstruction [76,107,108], reconstruction of color images [109], super-resolution [73], as well as to perform cross-modality image transformations such as virtual labeling of histological tissues [110], live cells [111], and to give brightfield image contrast to holographic images [64].

Here, we build upon these advances and present a novel deep learning-based holographic polarization microscope (DL-HPM) which can provide the retardation and orientation of birefringent specimen using a single phase-retrieved hologram that encodes one state of polarization (Fig. 4.1). This system requires only minor changes to the existing imaging hardware, i.e., the addition of a polarizer / analyzer set to a standard inline lensfree holographic microscope. Our framework uses SCPLM images as the ground truth to train a deep neural network (Fig. 4.1b), which uses this image data to learn how to transform the information encoded within a reconstructed hologram into an image that directly reveals the specimen's birefringence retardance and orientation. In addition to achieving a comparable image quality to the SCPLM images (Fig 4.1c), this deep learning-enabled lensfree microscopy method has a FOV of >20 mm^2 using a cost-effective optical design. The performance of DL-HPM is demonstrated using MSU and triamcinolone acetonide (TCA) crystal samples as well as a corn starch sample, matching the performance to SCPLM, both qualitatively and quantitatively.

We believe that the presented deep-learning based polarization microscopy approach could be widely used as a diagnostic tool in pathology and other fields that need to rapidly process and reveal the unique signatures of various birefringent crystals within complex specimen such as synovial fluid samples.

Part of this chapter has been published in:

Liu, T., de Haan, K., Bai, B., Rivenson, Y., Luo, Y., Wang, H., ... & Ozcan, A. (2020). Deep learning-based holographic polarization microscopy. ACS photonics, 7(11), pp.3023-3034.

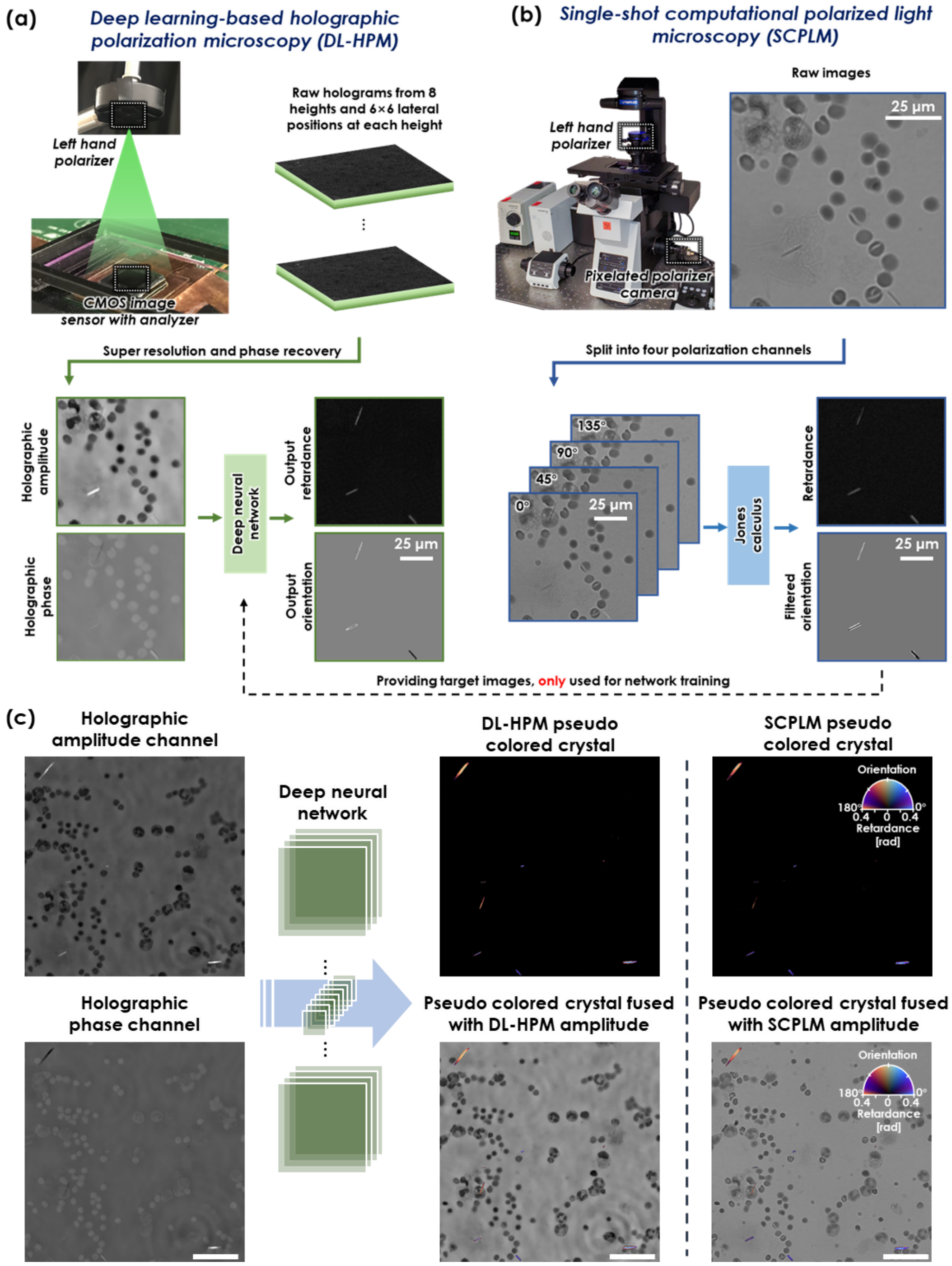


Figure 4.1 (a) Schematic for deep learning-based holographic polarization microscopy (DL-HPM). Raw holograms are collected using a lensfree holographic microscope with a customized polarizer and analyzer. A trained neural network is used to transform the reconstructed holographic amplitude and phase images into the birefringence retardance and orientation images. **(b) Schematic for single-shot computational polarized light microscopy (SCPLM).** Images are collected with a four-channel pixelated polarized camera under circularly polarized illumination. Birefringent retardance and orientation channels are computed using Jones calculus, and the amplitude image is

obtained by averaging the four polarization channels. SCPLM is used as the ground truth information channel, providing the network training target for DL-HPM. (c) **Blind testing of DL-HPM**. A new clinical sample (containing MSU crystals) collected from a de-identified patient is tested using DL-HPM. Birefringent samples are given a pseudo color using the same convention according to the compensated polarized light microscopy. Similar image quality was achieved compared to SCPLM images. Scale bar: 50 μm.

4.2 Results and discussion

We trained a deep neural network (see the Methods section) using 6 clinical samples containing MSU crystals, collected from 6 de-identified patients, to perform an image transformation, from an input holographic image (amplitude and phase) to the birefringence retardance and orientation images at the output of the network. The slides were all reviewed using CPLM (Olympus BX-51) by our clinical expert (JF) for the presence of MSU crystals. This analysis found that the majority of the birefringent crystals within these samples are needle shaped MSU crystals. Once trained, the neural network was blindly tested with 2 additional MSU slides from 2 new patients; Fig. 4.2(a) shows the blind output of the DL-HPM method in comparison to the SCPLM method. The birefringent crystals within the FOV are colored using a calibrated colormap according to the CPLM convention, where the background is left in grayscale to enhance the contrast. Fig. 4.2(b) further shows two representative zoomed-in regions for both single MSU crystals (within a blood cell) and a crystal cluster. These images demonstrate that our deep learning framework is capable of accurately locating the birefringent objects and giving them a high color contrast with respect to the non-birefringent cell background within the FOV.

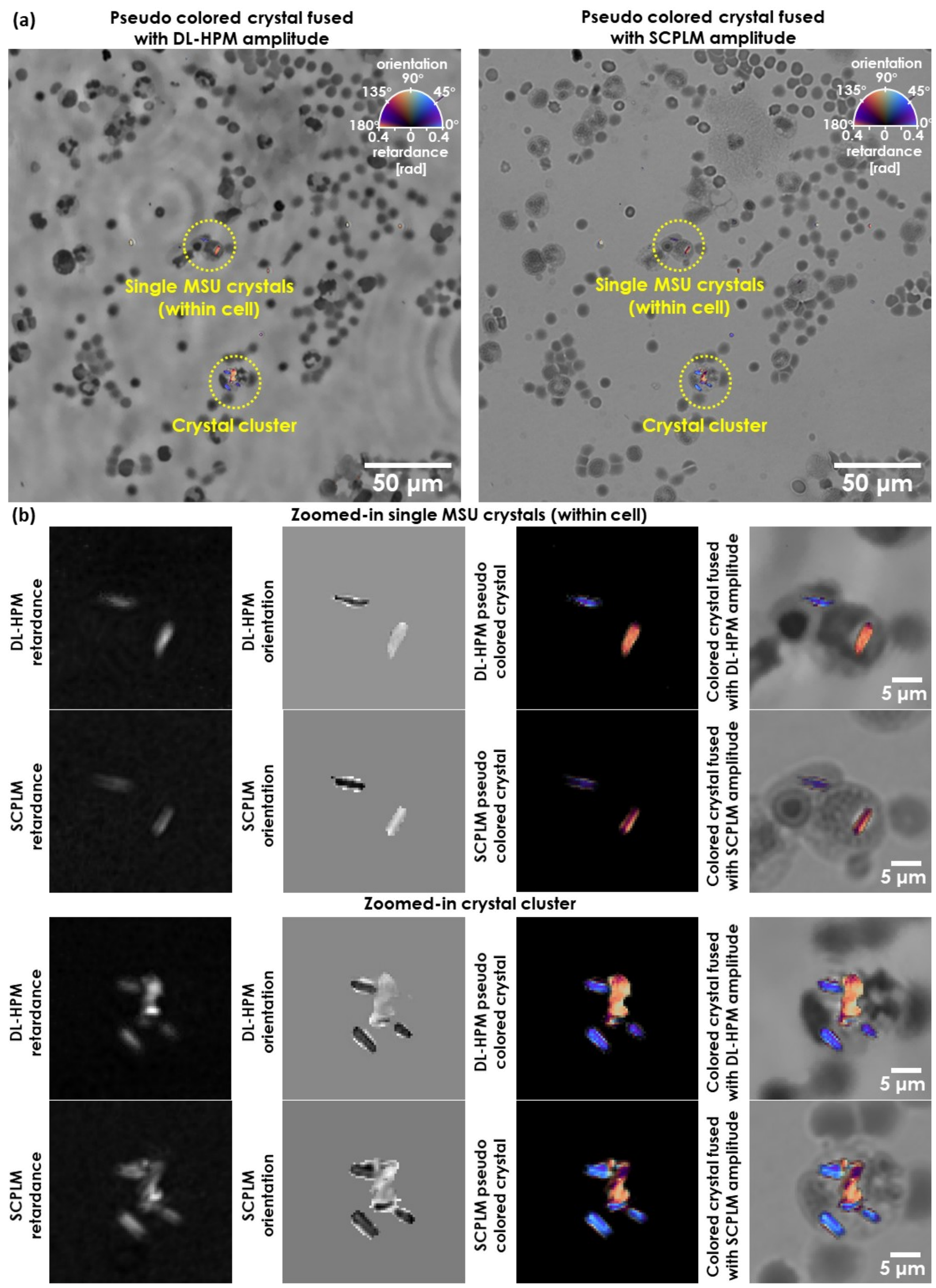


Figure 4.2 Imaging performance of DL-HPM. (a) Images generated using DL-HPM compared against co-registered images captured using SCPLM for a blindly tested MSU sample. The birefringent MSU crystals are colored according to CPLM convention after obtaining the retardance and orientation channels using each method. The pseudo-colored retardance and orientation information is also fused with the amplitude channel to show the high contrast against the non-birefringent cell background achieved by both methods. (b) Two different zoomed-in regions cropped

from (a).

To quantify the performance of our method, 3432 different birefringent objects were detected and analyzed. These objects were composed of individual MSU crystals (most common), MSU crystal clusters, or protein clusters (rare). We classified the birefringent objects according to their length and analyzed each of these categories separately. The classification was performed by first setting a 0.2 Rad threshold on the retardance channel in SCPLM target to convert it into a binary mask. Then, using these masks, connected component analysis was performed to classify each object into different length categories. For each detected birefringent object in the SCPLM, the same coordinates were also used to locate the corresponding birefringent object in the co-registered DL-HPM. The absolute retardance/orientation error was computed pixel-wise and averaged by the number of pixels for each detected birefringent object at its local region within a 5-pixel radius of any edge of the object.

The results of this quantitative analysis are reported in Fig. 4.3(a) with sample FOVs provided for visual comparison in Fig. 4.3(b). The minimum length of the crystals included in this analysis was selected to be 2 μm, representing an object with at least 5 pixels in length. Objects smaller than this threshold had insufficient resolution to assign a crystal type accurately. In total, 6 length categories were selected: 2-4 μm (1077 objects), 4-6 μm (582 objects), 6-8 μm (454 objects), 8-10 μm (466 objects), 10-20 μm (795 objects), and 20-50 μm (58 objects). The error was first computed and averaged pixel-wise for each detected birefringent object, and then averaged object-wise to obtain the statistics reported in Fig. 4.3(a). The overall object-wise averaged absolute error between the SCPLM results and the DL-HPM results was 0.047 Rad in the retardance channel and 0.135 Rad in the orientation channel, where the retardance and orientation channels range from 0 to $\pi/2$ and 0 to π, respectively. In order to further compare the performance of the DL-HPM with SCPLM, we plotted in Fig. 4.4 the cross-sections of the retardance and orientation channels for various crystals from the middle column of Fig. 4.3b in each length category (detailed error analysis is shown in Supporting Information). These results reveal that DL-HPM can quantitatively transform the holographic amplitude/phase information that is acquired at a single polarization state into birefringence retardance and orientation image channels, closely matching the results of SCPLM.

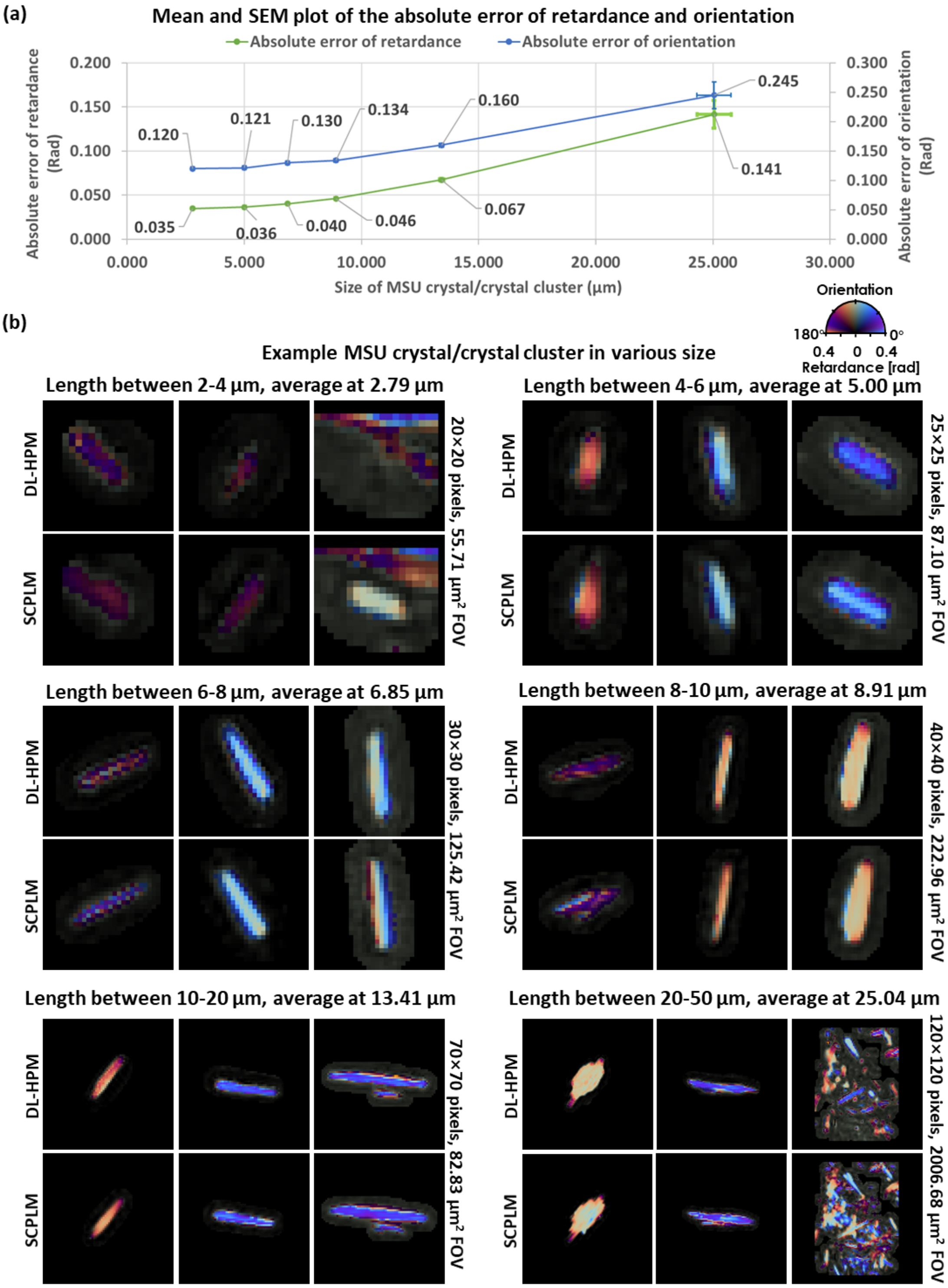


Figure 4.3 Quantitative performance of DL-HPM, compared against SCPLM results. (a) Mean and standard error of the mean (SEM) plots of the absolute error. DL-HPM achieves an overall object-wise averaged absolute error of 0.047 Rad in the retardance channel and 0.135 Rad in the orientation channel. (b) Sample birefringent objects. For each length category, the left image is the smallest object and the right is the largest, while the middle one has the median size.

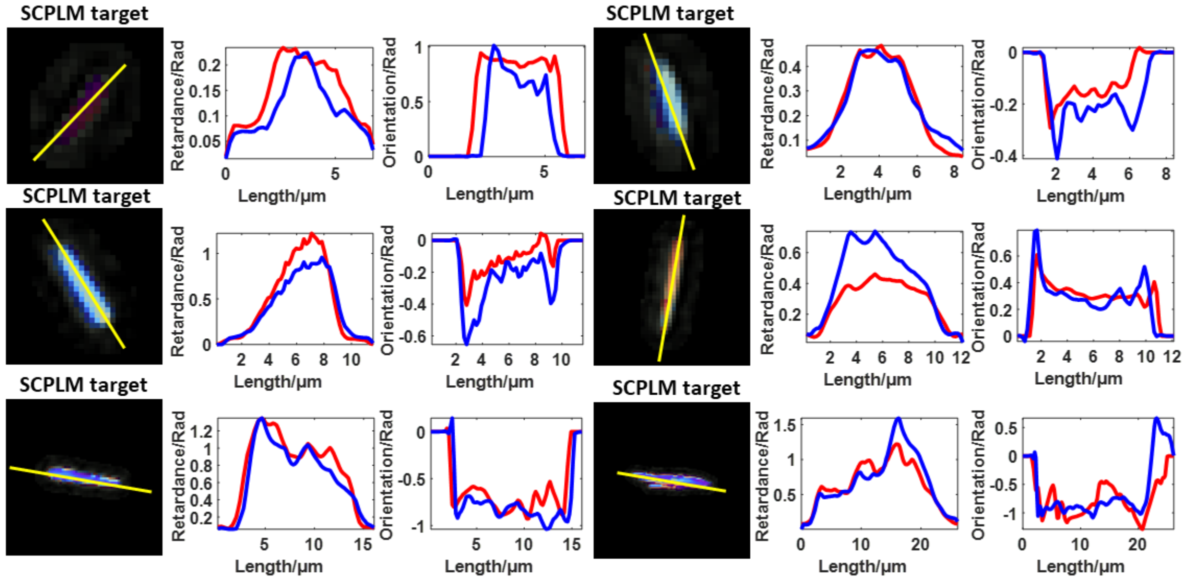


Figure 4.4 Cross section plots of the retardance and orientation channels obtained by DL-HPM output and SCPLM target. The blue curve represents the output of DL-HPM, while the red curve represents the target from SCPLM.

To further investigate the image transformation performed by the trained deep network, next we blindly tested it on *two new types of birefringent samples* that were never seen by the network during its training; for this purpose, we imaged TCA crystals, and corn starch samples (Fig. 4.5). The results revealed that DL-HPM can correctly identify birefringence in most of the TCA crystals and corn starch particles within the sample FOV. This is an indication that the presented DL-HPM method is learning a combination of the desired physical image transformation and a semantic segmentation which is related to morphological information of the samples. However, the network is observed to be less accurate when applied to these new types of samples than it is when tested upon samples of the same type that it was trained with.

Therefore, transfer learning was used to improve its generalization for new types of samples. As expected, once the same network is further trained using transfer learning with these new types of samples, its blind inference performance can be improved, which is illustrated in Fig. 4.6.

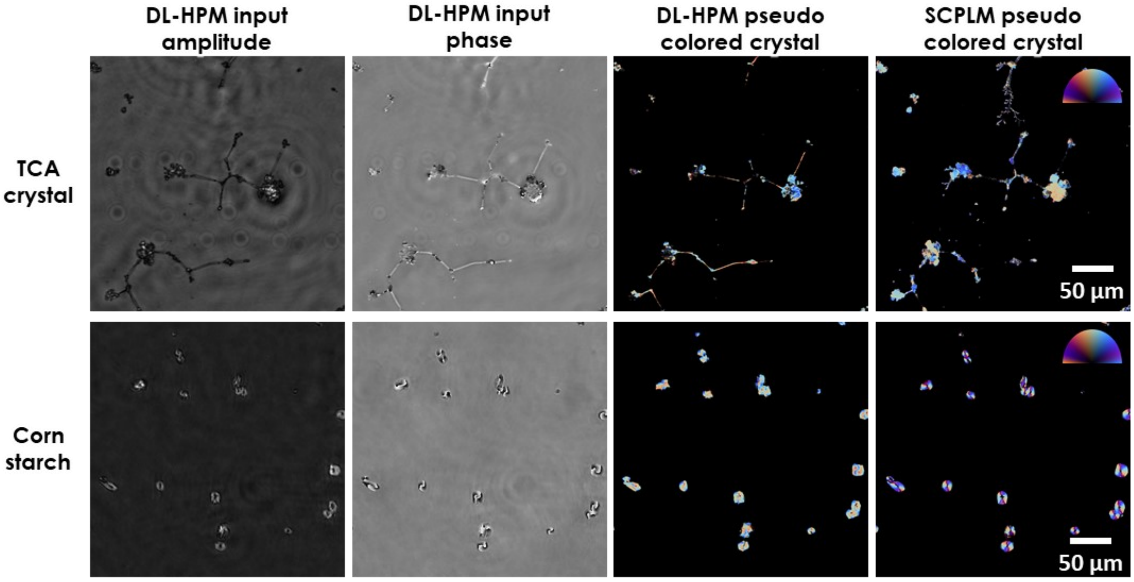


Figure 4.5 Imaging performance of DL-HPM on new types of birefringent samples. Visualization of birefringent TCA crystals and corn starch samples imaged using DL-HPM method, where the associated deep network is trained with only birefringent MSU crystals. Color bar: from left to top to right, represents π, π/2, 0 Rad in the orientation channel. Retardance is represented by the distance from the center of the color bar, ranging from 0 to 0.4.

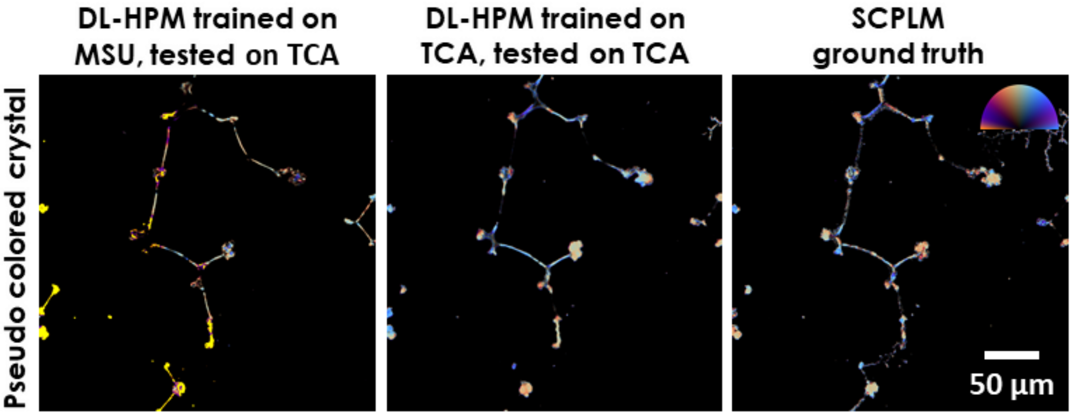


Figure 4.6 Comparison of two different deep neural networks for imaging TCA samples using

DL-HPM. The first network is trained with only MSU samples, and the second network is trained through transfer learning from the original MSU network with a training dataset containing TCA samples. Color bar: from left to top to right, represents π, π/2, 0 Rad in the orientation channel. Retardance is represented by the distance from the center of the color bar, ranging from 0 to 0.4.

Our results have qualitatively and quantitatively demonstrated the effectiveness of the presented framework using multiple types of samples, also illustrating the capability of the networks to generalize from one sample type to another. However, the black box nature of deep neural networks often makes it challenging to determine how the transformation is actually performed. In this section, we perform an ablation study aiming to partially reveal the physical interpretation of the deep neural network, and demonstrate that the network is learning to perform crystal segmentation based on both the morphological information and the physical relationship between the holographic amplitude/phase information and the birefringent retardance/orientation channels. For this analysis, we trained two additional networks using the MSU image dataset: 1. using only the holographic *amplitude channel* as the input to the neural network to blindly perform the retardance/orientation inference; and 2. using only the holographic *phase channel* as the input to the neural network to blindly perform the retardance/orientation inference. Examples of the blind inference performance of these trained networks are shown in Fig. 4.7.

In general, using only the amplitude or only the phase channel, as opposed to using both channels together, significantly degrades the inference performance of the network. The amplitude only network tends to accurately predict the crystals but generates images with significant error in the orientation channel, whereas the phase only network tends to predict the locations of the crystals in the retardance channel less accurately. One possible explanation for this observation is that the amplitude only network is learning morphological information to locate the crystals, as the holographic optical system was designed to introduce intensity contrast for birefringent samples [98], and the additional phase information is required to fully reconstruct the orientation channel. Hence, both the amplitude and phase information channels are essential to accurately infer the retardance and orientation of birefringent samples. This conclusion is also supported by analyzing the formulation of HPM, which will be discussed next.

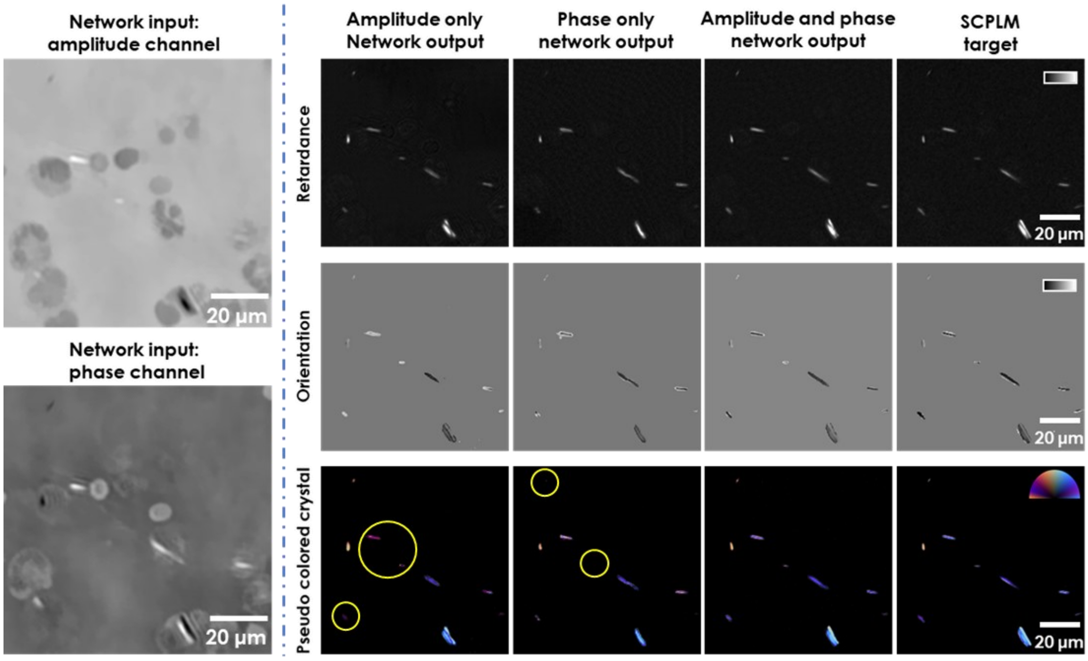


Figure 4.7 DL-HPM reconstruction results using different input channels. The performance of DL-HPM in general degraded when reducing the input channel to amplitude only or phase only information. Color bar for retardance, from black to white, represents 0 to 1 Rad. Color bar for orientation, from black to white, represents -π/2 to π/2. Color bar for pseudo-colored crystals, from left to top to right, represents π, π/2, 0 Rad in the orientation channel. Retardance is represented by the distance from the center of the color bar, ranging from 0 to 0.4. The circled crystals in the amplitude only network show that the network is capable of reconstructing the crystal features in the correct location, but the color is inaccurate due to the errors in the orientation channel. The circled crystals in phase only network are either entirely missing or have a retardance that is too low, rejected by the pseudo coloring scheme.

The evolution of the polarization state in our optical setup (Fig. 4.1a) can be analyzed using

Jones calculus (see Methods section), where for *each pixel* of the reconstructed holographic amplitude and phase images, we can write:

$$\begin{cases} A_{\text{recon}} = \dfrac{\sqrt{a^2+b^2}}{\left|\cos\beta - \sin\beta\right|} \\ \varphi_{\text{recon}} = \operatorname{atan2}\left(b,a\right) - \dfrac{\eta}{2} - \dfrac{\pi}{2} - \operatorname{atan2}\left[0, \cos\beta - \sin\beta\right] \end{cases} \tag{15}$$

where atan2(y, x) is the four-quadrant inverse tangent function for point (x, y), A_{recon} is the normalized reconstructed amplitude, φ_{recon} is the normalized reconstructed phase (with zero phase in the background), β represents the orientation of the linear polarizer with respect to the x-axis, and a and b are defined as:

$$\begin{cases} a = -\cos\beta\sin\eta\sin^2\theta + \sin\beta\sin\eta\cos^2\theta \\ \quad +\cos\beta\cos\theta\sin\theta + \sin\beta\cos\theta\sin\theta \\ \quad -\cos\beta\cos\eta\cos\theta\sin\theta - \sin\beta\cos\eta\cos\theta\sin\theta \\ b = \cos\beta\cos^2\theta - \sin\beta\sin^2\theta \\ \quad +\cos\beta\cos\eta\sin^2\theta - \sin\beta\cos\eta\cos^2\theta \\ \quad -\cos\beta\sin\eta\cos\theta\sin\theta - \sin\beta\sin\eta\cos\theta\sin\theta \end{cases} \tag{16}$$

θ represents the orientation of the fast axis of the sample with respect to the x-axis, and η represents the relative phase retardance.

Unlike the SCPLM method, where the retardance, η, and the orientation, θ, are encoded in symmetrical equations with a straightforward analytical solution (detailed in the Supporting Information), Equations 15-16 encode the birefringence information in a much more convoluted form. Because of the experimental challenges in obtaining an accurate estimate of β as well as the potential phase wrapping related issues, independently solving Equations 15-16 in a pixel-by-pixel manner could result in errors or spatial inconsistencies/artifacts at the output retardance and orientation images. Hence, it elevates the need for a more advanced solution and a robust method such as a deep neural network, which is trained to perform an image-to-image transformation by making use of all the information from multiple pixels within a FOV *simultaneously*. Stated differently, through image data, the deep neural network learns to solve Equations 15-16 over an input FOV, where all the pixels within the complex-valued input image (phase and amplitude) are simultaneously processed to generate the desired output image channels, i.e., the retardance (η) and the orientation (θ) images.

4.3 Conclusions

We presented a deep learning-enabled holographic polarization microscope. This framework is advantageous as it only requires the measurement of a single polarization state which can be generated using a simple optical setup, and is capable of accurately reconstructing the quantitative birefringent retardance and orientation information of the specimen. These information channels can dramatically simplify the automatic detection, counting, and classification of birefringent objects within complex media. After necessary regulatory testing and approvals, our method can be the basis of a rapid point-of-care crystal detection and analysis instrument with automated crystal identification and classification capabilities, which could significantly simplify the clinical procedures used to diagnose diseases related to birefringent crystals, such as gout and pseudogout. In addition, we also expect our DL-HPM method to benefit a wider range of technical fields for both research and industrial applications, e.g., surface defect detection and air quality monitoring [112] .

4.4 Materials and Methods

Lensfree polarization imaging setup

The presented DL-HPM system utilizes a customized lensfree holographic polarization microscope to capture the input images (Fig. 4.1a). This microscope is able to generate quantitative phase images as well as introduce an intensity contrast to birefringent objects (though it is unable to differentiate high absorbance non-birefringent objects [98]). The microscope uses a laser source filtered by an acousto-optic tunable filter (AOTF) for illumination at 530 nm (~2.5 nm bandwidth). The raw holograms were collected using a CMOS image sensor (IMX 081, Sony, 1.12μm pixel size) at 8 sample to sensor heights, which were used for multi-height phase recovery [34]. A set of low-resolution holograms were captured at 6×6 lateral positions, which were used for pixel-super resolution. Using these images, a high-resolution holographic image was reconstructed and subsequently numerically back propagated [69] to the sample plane using an auto-focusing algorithm [38]. Finally, the reconstructed hologram was normalized to obtain an average background amplitude of 1, and have an average background phase of 0. These normalized images were then passed through the neural network. Details of the holographic image reconstruction techniques including free space propagation, multi-height phase recovery, super resolution, and auto-focusing are presented in the Supporting Information.

To enable detection of the birefringence within the sample, a left-hand polarizer and a customized analyzer were added in the holographic imaging system [98]. Unlike traditional polarization microscopes, where a second circular polarizer (i.e. right-hand polarizer) can be used as the analyzer, holographic imaging systems require background light to form an interference pattern, and the direct use of another circular polarizer would completely reject the background light. Therefore, we designed the analyzer to use a $\lambda/4$ retarder film (75 μm thickness, Edmund Optics), and a linear polarizer (180 μm thickness, Edmund Optics), having the fast axis of the $\lambda/4$ retarder oriented to 25° with respect to the linear polarizer, creating a holographic polarization microscope. These films were affixed directly to the CMOS image sensor using an ultraviolet (UV)-curable adhesive (NOA 68, Norland Products, Cranbury, NJ) as shown in Fig. 4.8b.

Polarization encoding in the holographic imaging system

In the analysis of our holographic imaging system, we assume that the sample, polarizer, and analyzer are thin and have negligible gaps between them. We further assume that the light diffracts from the analyzer onto the image sensor after being converted to linearly polarized light by the last layer of the analyzer. Therefore, after its reconstruction, the hologram becomes in-focus at the sample plane (the thicknesses of both the sample and the analyzer are assumed to be negligible).

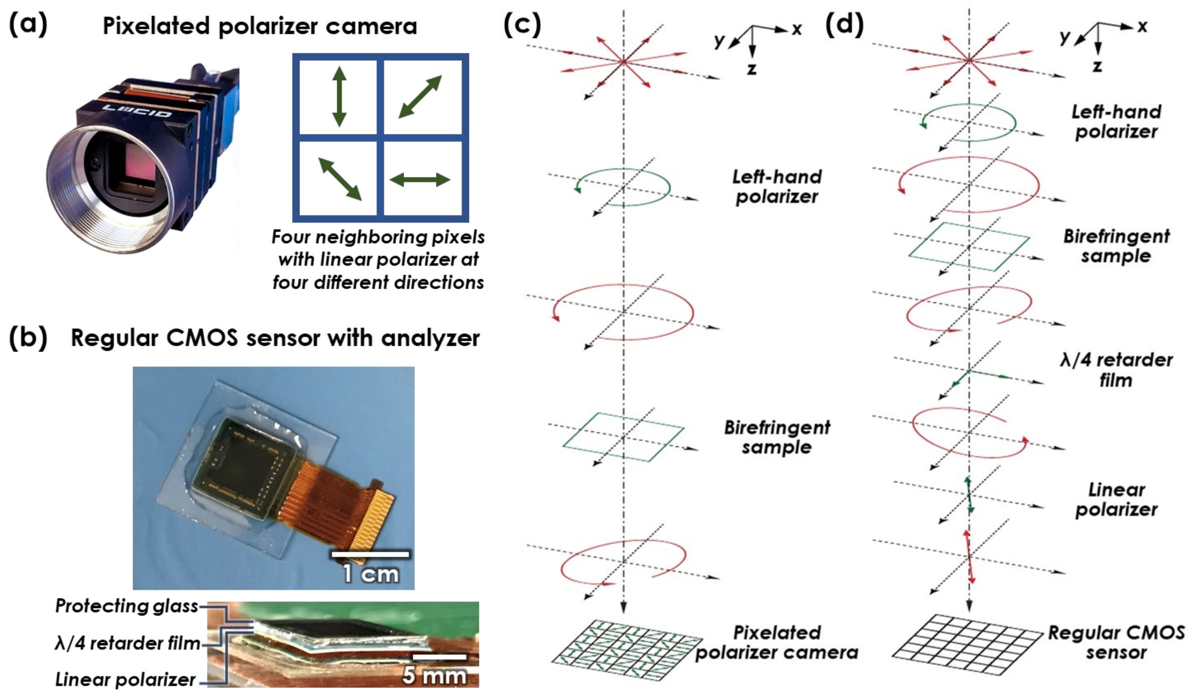


Figure 4.8 Images sensor photos and polarization designs of the SCPLM and DL-HPM systems. (a) Photo of the four-channel pixelated polarizer camera. This camera is used in SCPLM. When illuminated with circularly polarized light, four channels with different polarization states are acquired using a single image. (b) Photo of the CMOS image sensor with a customized analyzer. This imager is used for DL-HPM. The analyzer film is directly bound to the CMOS image sensor, allowing a certain amount of background light to form the hologram, while also providing one polarization channel to sense the birefringent sample. (c) Polarization design for SCPLM. The red plot represents the light polarization state at each plane. The green plot represents an optical component that changes the polarization state. (d) Polarization design for DL-HPM.

The evolution of the polarization in our imaging system can be analyzed using Jones calculus [113]. The light field in the presented framework was designed to be transmitted through a left-hand polarizer, birefringent sample, $\lambda/4$ retardation plate and a linear polarizer (Fig 4.2d). Each of these optical components can be formulated as:

I. Input left-hand circularly polarized (LHCP) light:

$$\mathbf{E}_{\text{in}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -i \end{bmatrix} \tag{17}$$

where $i^2 = -1$, and LHCP is defined *from the point of view of the source*.

II. Birefringent sample:

$$\mathbf{M}_{\text{sample}} = e^{-\frac{i\eta}{2}}\begin{bmatrix} \cos^2\theta + e^{i\eta}\sin^2\theta & \left(1-e^{i\eta}\right)\cos\theta\sin\theta \\ \left(1-e^{i\eta}\right)\cos\theta\sin\theta & \sin^2\theta + e^{i\eta}\cos^2\theta \end{bmatrix} \tag{18}$$

where θ represents the orientation of the fast axis of the sample with respect to the x-axis, and η represents the relative phase retardance.

III. $\lambda/4$ retarder:

$$\mathbf{M}_{\text{retarder}} = e^{-\frac{i\pi}{4}}\begin{bmatrix} \cos^2\alpha + i\sin^2\alpha & (1-i)\cos\alpha\sin\alpha \\ (1-i)\cos\alpha\sin\alpha & \sin^2\alpha + i\cos^2\alpha \end{bmatrix} \tag{19}$$

where α represents the orientation of the fast axis of the $\lambda/4$ retarder with respect to the x-axis.

IV. Linear polarizer:

$$\mathbf{M}_{\text{linear}} = \begin{bmatrix} \cos^2\beta & \cos\beta\sin\beta \\ \cos\beta\sin\beta & \sin^2\beta \end{bmatrix} \tag{20}$$

where β represents the orientation of the linear polarizer with respect to the x-axis.

The output light field can then be expressed as:

$$\mathbf{E}_{\text{out}} = \mathbf{M}_{\text{linear}}\mathbf{M}_{\text{retarder}}\mathbf{M}_{\text{sample}}\mathbf{E}_{\text{in}} \tag{21}$$

Equations 15-16 reported in the Discussion section are obtained by rearranging Equation 21, and applying a background normalization step (detailed in the Supporting Information).

Dataset preparation

The neural networks were trained using image pairs captured using both SCLPM and the holographic imaging systems. We used 6 clinical MSU samples for training, and 2 additional MSU samples for testing. To ensure that the network training can generalize to new samples, the slides used to train the neural network were chosen to have different concentrations of MSU crystals (example FOVs of each slide are shown in Supporting Information Fig. S4.3). In addition, a single TCA sample was used for both training and testing (where blind testing was performed on new regions); similarly, a single starch sample was used for blind testing.

In order to train the neural network to learn the image transformation from a lensfree holographic imaging modality to a lens-based SCPLM system, an accurately co-registered training dataset is required. This co-registration begins by bicubic down-sampling the target polarization images by a factor of 0.345/0.373 (obtained with the benchtop microscope) to match the pixel size of the holographic microscope; the ground truth images were created using SCPLM with an effective pixel size of 0.345 μm, while the holographic images used as the network input have an effective pixel size of 0.373 μm. Next, a rough matching between the two sets of images is obtained by finding the overlapping area with maximum correlation between the amplitude channels of the images. Once the images have been roughly aligned, global matching is performed by applying an affine transformation calculated using MATLAB's multimodal image registration framework [114]. This framework extracts features from the amplitude channel of the images and matches them with the affine matrix. Next, large fields of view were cropped, and matched to each other using an elastic pyramidal registration algorithm, which allows for pixel level matching [115] based upon the amplitude channel of the images. As discussed above, these holograms were normalized to have an average background amplitude of 1, and an average background phase of 0, before being passed through the neural network. This normalization step helps the neural network to be applied more consistently to different samples. The orientation of areas of the polarization images without any birefringence are set to zero to eliminate noise in the labels. This is done by setting the orientation value of any pixel (below a threshold) in the corresponding retardance channel to zero.

Neural network

A Generative Adversarial Network (GAN) framework was used to perform the image transformation. In addition to this GAN loss, a mean absolute error (L_1) loss was used to ensure that the transformation is accurate, and a total variation (TV) loss is used as a regularization term. GANs use two separate networks for their operation. A generator network ($G(\cdot)$) is used to transform the holographic images into their polarization counterparts. A second network known as the discriminator ($D(\cdot)$) is used to discriminate between the ground truth images (z) and the generated images ($G(x)$). The two networks learn from one another, with the generator gradually learning how to create images that match the feature distribution of the target dataset, while the discriminator gets better at distinguishing between ground truth images and their generated counterparts.

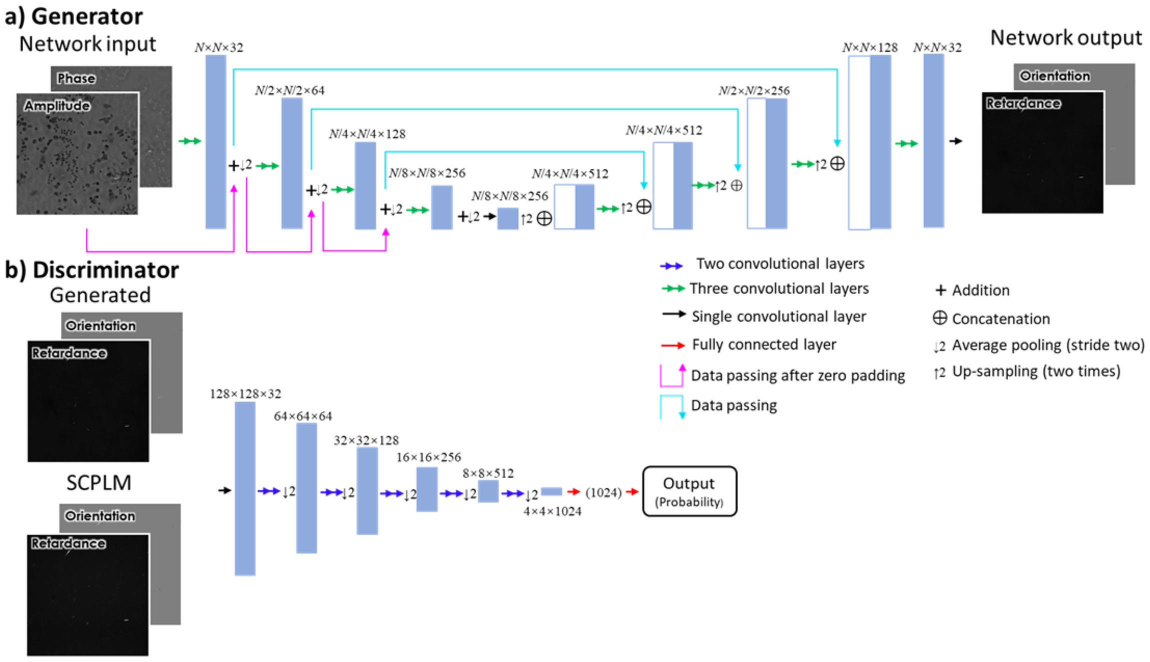


Figure 4.9 Network architecture. (a) Diagram of the generator portion of the network. (b) Diagram of the discriminator portion of the network.

The overall loss function can be described as:

$$l_{\text{generator}} = \mathrm{L}_1\left\{\mathrm{G}(x),z\right\} + \lambda_1 \times \mathrm{TV}\left\{\mathrm{G}(x)\right\} + \lambda_2 \times \left(1-\mathrm{D}\left(\mathrm{G}(x)\right)\right)^2 \tag{22}$$

where λ_1 and λ_2 are constants used to balance the various terms of the loss function. The L_1 loss, which ensures that the transformation performed by the network is accurate, was balanced to make up ~25% of the total loss function, while the total variation loss makes up ~0.5% of the overall loss function. The L_1 loss is defined as:

$$\mathrm{L}_1\left\{\mathrm{G}(x),z\right\} = \frac{1}{M \times N \times K}\sum_k \sum_i \sum_j \left| z_{i,j,k} - \mathrm{G}(x)_{i,j,k} \right| \tag{23}$$

where K is the number of image channels ($k=1$ represents the retardance channel, and $k=2$ represents the orientation channel), M and N are the number of pixels on each axis, and i and j represent the pixel indices of the image. The total variation loss, used to regularize the loss function and reduce noise is defined as:

$$\mathrm{TV}\left(\mathrm{G}(x)\right)=\frac{1}{M\times N\times K}\sum_{k}\sum_{i}\sum_{j}\left|\mathrm{G}(x)_{i+1,j,k}-\mathrm{G}(x)_{i,j,k}\right|+\left|\mathrm{G}(x)_{i,j+1,k}-\mathrm{G}(x)_{i,j,k}\right| \quad (24)$$

In order to train the discriminator a separate loss function is used, defined as:

$$l_{\text{discriminator}}=\mathrm{D}\left(\mathrm{G}(x)\right)^{2}+\left(1-\mathrm{D}(z)\right)^{2} \quad (25)$$

Depending on the exact dataset being used and the application of interest, the relative weights of the different portions of these loss functions, or even the makeup of the loss function itself can be changed.

The generator network uses the U-net architecture [40], as shown in Fig. 4.9(a). The U-net architecture is used, as this architecture has been shown to be highly effective for biological applications, as it can learn to transform features at multiple different scales. This U-net begins with a convolutional layer increasing the number of channels to 32, and is made up of four “down-blocks” followed by four “up-blocks”. Each down-block consists of three convolutional layers, which together double the number of channels. These layers are followed by an average pooling layer with a kernel size and stride of 2. After these down-blocks, an equal number of up-blocks are applied. The up-blocks begin by bilinear up-sampling the images and similarly apply three convolutional layers, and reduce the number of channels by a factor of four. Between the two sets of blocks, skip connections are added. The skip connections allow small scale data to pass through the network, avoiding the effects of the down-sampling by concatenating the output of each down-block with the input to each up-block. Following these blocks, a convolutional layer reduces the number of channels to two, which match the two channels of the polarization images.

The discriminator network (Fig. 4.9(b)) receives the generated images or the SCPLM images, and attempts to distinguish between the two. The discriminator is first made up of a convolutional layer which increases the number of channels from 3 to 32. This is followed by five blocks, each made up of two convolutional layers, the second of which doubling the number of channels and using a stride of two. Following these five blocks are two fully connected layers, which reduce the image to a single number which can have a sigmoid function applied to it.

Each convolutional layer uses a kernel size of 3×3 and is followed by the leaky rectified linear unit (LeakyReLU) activation function which is defined as:

$$\text{LeakyReLU} = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \tag{26}$$

In the training phase, we used the adaptable movement estimation (Adam) optimizer, with a learning rate of 10^{-4} for the generator, and 10^{-5} for the discriminator. The network begins with the generator being trained 7 times for each training of the discriminator, with this ratio being reduced by 1 every 4000 iterations down to a minimum of 3. The network was trained for 30,000 iterations of the discriminator, taking 5 hours using a standard consumer GPI. This training was performed using a single 2080 Ti (Nvidia), with Python version 3.6.0 and TensorFlow Version 1.11.0. MATLAB version R2018a was used for preprocessing.

We also want to emphasize that, though our network model is sample type-specific, training a network model is a one-time effort. For each novel application scenario, a new model is required, which must be generated with the similar measurement and training procedures. However, once the new model is determined, it is ready to be distributed to a large number of end users, without requiring extra cost or resources.

4.5 Appendix

Holographic image reconstruction

Free space propagation. Free space wave propagation [69] is performed using the following equation:

$$\mathrm{FT}\{U(x,y;z)\} = \mathrm{FT}\{U(x,y;0)\}\cdot H(f_x,f_y;z) \tag{27}$$

where FT{·} is the 2D Fourier transform operation, $U(x, y; z)$ is the light field at an axial distance z, and $H(f_x, f_y; z)$ is defined as:

$$H(f_x,f_y;z)=\begin{cases} 0, & \left(\frac{\lambda f_x}{n}\right)^2+\left(\frac{\lambda f_y}{n}\right)^2>1 \\ \exp\left[j2\pi\frac{n}{\lambda}z\sqrt{1-\left(\frac{\lambda f_x}{n}\right)^2-\left(\frac{\lambda f_y}{n}\right)^2}\right], & \text{o.w.} \end{cases} \tag{28}$$

where f_x, f_y are the spatial frequencies along x and y directions, respectively, λ is the wavelength, and n is the refractive index of the medium.

Autofocusing. To accurately estimate the distance between the image sensor and the sample, an edge sparsity criterion (Tamura of the gradient (ToG)) based autofocusing is used [38]. First, we propagate the unfocused hologram to a series of z distances. At each distance, the ToG is computed, and the distance with the highest ToG is treated as the focusing distance. If the unfocused hologram is an amplitude only hologram (e.g. raw hologram captured using the image sensor), the propagation algorithm assumes zero phase by default.

Multi-height phase recovery. To eliminate twin-image related artifacts we used an iterative multi-height phase recovery algorithm [34,36,116]. Raw holograms from 8 sample to sensor distances with ~10 µm step size were collected. The hologram at the 1st height was padded with a zero-phase channel, and then propagated to the 2nd height. After the axial propagation, the amplitude channel values are averaged with the measured amplitude channel at the 2nd height, where the propagated phase channel was kept unchanged. This process is continued until it has been propagated to the 8th height, then back to the 1st height, which defines one iteration. The missing phase channel of a given hologram is typically recovered after 10-30 iterations.

Shift-and-add pixel super resolution. A super resolution algorithm [117] was used to improve the resolution of each hologram prior to the multi-height phase recovery step. Thirty-six low-resolution raw holograms at different lateral positions (6x6 with ~0.3733 μm spacing along the x and y directions) were collected at each sample-to-sensor distance to generate these high-resolution holograms. To accurately estimate the relative distance between the low-resolution holograms, a correlation-based method was used [35]. With the distance of these shifts being accurately obtained, the shift-and-add super resolution algorithm was used to fuse the low-resolution holograms into a high-resolution corresponding hologram.

SCPLM mathematical model derivation

The design of the single-shot computational polarized light microscopy (SCPLM) setup [97] is demonstrated in the main text Figure 7(c). A birefringent sample of interest is illuminated using left-hand circularly polarized (LHCP) light, which is obtained by placing a left-hand circular polarizer behind an LED illumination source. The transmitted light is recorded using a pixelated polarizer camera. As shown in Figure 7(a), every four pixels of the camera sensor utilize four different directional polarizing filters (0°, 90°, 45°, and 135°) and form a repeated pattern across the sensor.

The mathematical model of the SCPLM can be formulated by Jones calculus, where the optical components can be expressed as:

I. Input left-hand circularly polarized (LHCP) light:

$$\mathbf{E}_{\text{in}} = \frac{E_0}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix} \tag{29}$$

where $i^2 = -1$, E_0 is the amplitude of the incident LHCP light, and LHCP is defined from the point of view of the receiver.

II. Birefringent sample:

$$\mathbf{M}_{\text{sample}} = \begin{bmatrix} \cos\frac{\eta}{2} + i\sin\frac{\eta}{2}\cos 2\theta & i\sin\frac{\eta}{2}\sin 2\theta \\ i\sin\frac{\eta}{2}\sin 2\theta & \cos\frac{\eta}{2} - i\sin\frac{\eta}{2}\cos 2\theta \end{bmatrix} \tag{30}$$

where η and θ represent the retardance magnitude and the slow axis orientation of the birefringent object. Both η and θ are spatially varying.

III. Detector on the pixelated image sensor:

$$\mathbf{M}_{\text{detector}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \text{ or } \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{31}$$

for four different directional polarizing filters at 0°, 45°, 90°, and 135°, respectively.

The output light field can be expressed as:

$$\mathbf{E}_{\text{out}} = \mathbf{M}_{\text{detector}}\mathbf{M}_{\text{sample}}\mathbf{E}_{\text{LHCP}} \tag{32}$$

Consider a pixel with 0°-orientated polarizing filter, the corresponding light field can be formulated as:

$$\mathbf{E}_{\text{out}}^{0} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \cos\frac{\eta}{2} + i\sin\frac{\eta}{2}\cos 2\theta & i\sin\frac{\eta}{2}\sin 2\theta \\ i\sin\frac{\eta}{2}\sin 2\theta & \cos\frac{\eta}{2} - i\sin\frac{\eta}{2}\cos 2\theta \end{bmatrix} \cdot \frac{E_0}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix} \tag{33}$$

Then the intensity value recorded by this 0°-orientated pixel can be written as:

$$I_0 \propto \left|\mathbf{E}_{\text{out}}^{0}\right|^2 = \mathbf{E}_{\text{out}}^{0} \cdot \mathbf{E}_{\text{out}}^{0\,*} \tag{34}$$

Combining Equation 33 and 34, we get:

$$I_0 = \frac{1}{2} I_{\max}(1 - \sin\eta \sin 2\theta) \tag{35}$$

where $I_{\max}$ is a constant related to the intensity of the illumination light. Following the same steps, the intensity values captured by other three different orientated pixels of the image sensor chip can be expressed as:

$$
\begin{aligned}
I_{45} &= \frac{1}{2} I_{\max} (1 + \sin\eta \cos 2\theta), \\
I_{90} &= \frac{1}{2} I_{\max} (1 + \sin\eta \sin 2\theta), \\
I_{135} &= \frac{1}{2} I_{\max} (1 - \sin\eta \cos 2\theta)
\end{aligned}
\tag{36}
$$

Gathering Equation 35 and 36, two auxiliary variables *A* and *B* can be defined as:

$$
\begin{aligned}
A &= \frac{I_{90} - I_0}{I_{90} + I_0} = \sin\eta \sin 2\theta \\
B &= \frac{I_{45} - I_{135}}{I_{45} + I_{135}} = \sin\eta \cos 2\theta
\end{aligned}
\tag{37}
$$

Finally, the retardance magnitude and the slow axis orientation of the birefringent object can be reconstructed as:

$$
\begin{aligned}
\eta &= \arcsin \sqrt{A^2 + B^2} \\
\theta &= \frac{1}{2} \arctan \frac{A}{B}
\end{aligned}
\tag{38}
$$

DL-HPM mathematical model

The mathematical model of the deep learning-based holographic polarization microscope (DL-HPM) which is described by Equation 21 in the main text can be expanded as:

$$
\begin{aligned}
\mathbf{E}_{\text{out}} = & \begin{bmatrix} \cos^2\beta & \cos\beta\sin\beta \\ \cos\beta\sin\beta & \sin^2\beta \end{bmatrix} e^{-\frac{i\pi}{4}} \begin{bmatrix} \cos^2\alpha + i\sin^2\alpha & (1-i)\cos\alpha\sin\alpha \\ (1-i)\cos\alpha\sin\alpha & \sin^2\alpha + i\cos^2\alpha \end{bmatrix} \\
& e^{-\frac{i\eta}{2}} \begin{bmatrix} \cos^2\theta + e^{i\eta}\sin^2\theta & \left(1-e^{i\eta}\right)\cos\theta\sin\theta \\ \left(1-e^{i\eta}\right)\cos\theta\sin\theta & \sin^2\theta + e^{i\eta}\cos^2\theta \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}
\end{aligned}
\tag{39}
$$

where the illumination light is normalized to have an amplitude of 1, LHCP is defined from the point of view of the source, *α* represents the orientation of the fast axis of the *λ*/4 retarder with respect to the x-axis, β represents the orientation of the linear polarizer with respect to the x-axis, θ represents the orientation of the fast axis of the sample with respect to the x-axis and η represents the relative phase retardance.

In our optical design, we selected α to be 90° without a loss of generality, and therefore we have:

$$\begin{aligned}\mathbf{E}_{\text{out}} = & \begin{bmatrix} \cos^2\beta & \cos\beta\sin\beta \\ \cos\beta\sin\beta & \sin^2\beta \end{bmatrix} e^{-\frac{i\pi}{4}} \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix} e^{-\frac{i\eta}{2}} \cdot \\ & \begin{bmatrix} \cos^2\theta + e^{i\eta}\sin^2\theta & \left(1-e^{i\eta}\right)\cos\theta\sin\theta \\ \left(1-e^{i\eta}\right)\cos\theta\sin\theta & \sin^2\theta + e^{i\eta}\cos^2\theta \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix}\end{aligned} \tag{40}$$

Then, using the terms before the linear polarizer, we have:

$$\mathbf{E}_{\text{before\_linear}} = \frac{1}{\sqrt{2}} e^{-\frac{i\pi}{4}} e^{-\frac{i\eta}{2}} \begin{bmatrix} i\cos^2\theta + ie^{i\eta}\sin^2\theta + \left(1-e^{i\eta}\right)\cos\theta\sin\theta \\ \left(1-e^{i\eta}\right)\cos\theta\sin\theta - i\sin^2\theta - ie^{i\eta}\cos^2\theta \end{bmatrix} \tag{41}$$

To further simplify the equation, we use Euler's formula:

$$\mathbf{E}_{\text{before\_linear}} = \frac{1}{\sqrt{2}} e^{-\frac{i\pi}{4}} e^{-\frac{i\eta}{2}} E_{\text{before\_linear\_x}} \cdot \mathbf{x} + \frac{1}{\sqrt{2}} e^{-\frac{i\pi}{4}} e^{-\frac{i\eta}{2}} E_{\text{before\_linear\_y}} \cdot \mathbf{y} \tag{42}$$

where **x** and **y** are unit vectors pointing toward the positive x and y axis directions, respectively, and $E_{\text{before\_linear\_x}}$ and $E_{\text{before\_linear\_y}}$ are the x and y components of $\mathbf{E}_{\text{before\_linear}}$, defined as:

$$\begin{cases} E_{\text{before\_linear\_x}} = i\cos^2\theta + \left(i\cos\eta - \sin\eta\right)\sin^2\theta + \left(1-\cos\eta - i\sin\eta\right)\cos\theta\sin\theta \\ E_{\text{before\_linear\_y}} = \left(1-\cos\eta - i\sin\eta\right)\cos\theta\sin\theta - i\sin^2\theta - \left(i\cos\eta - \sin\eta\right)\cos^2\theta \end{cases} \tag{43}$$

Since the final polarizer is a linear polarizer, the output field can be written as:

$$\mathbf{E}_{\text{out}} = \frac{1}{\sqrt{2}} e^{-\frac{i\pi}{4}} e^{-\frac{i\eta}{2}} \left(\cos\beta E_{\text{before\_linear\_x}} + \sin\beta E_{\text{before\_linear\_y}}\right) \cdot \mathbf{k} \tag{44}$$

where **k** is a unit vector pointing toward the direction with β degrees with respect to the positive x-axis.

Based on Equations 43-44, we get:

$$\mathbf{E}_{\text{out}} = \frac{1}{\sqrt{2}} e^{-\frac{i\pi}{4}} e^{-\frac{i\eta}{2}} \left(a + ib\right) \cdot \mathbf{k} \tag{45}$$

where a and b are both real and functions of θ and η that are defined by:

$$\begin{cases} a = -\cos\beta\sin\eta\sin^2\theta + \sin\beta\sin\eta\cos^2\theta \\ \quad + \cos\beta\cos\theta\sin\theta + \sin\beta\cos\theta\sin\theta \\ \quad - \cos\beta\cos\eta\cos\theta\sin\theta - \sin\beta\cos\eta\cos\theta\sin\theta \\ b = \cos\beta\cos^2\theta - \sin\beta\sin^2\theta \\ \quad + \cos\beta\cos\eta\sin^2\theta - \sin\beta\cos\eta\cos^2\theta \\ \quad - \cos\beta\sin\eta\cos\theta\sin\theta - \sin\beta\sin\eta\cos\theta\sin\theta \end{cases} \tag{46}$$

Then, the amplitude and phase of the output field can be written as:

$$\begin{cases} A_{\text{out}} = \dfrac{\sqrt{2}}{2}\sqrt{a^2 + b^2} \\ \varphi_{\text{out}} = \operatorname{atan2}\left(b, a\right) - \dfrac{\eta}{2} - \dfrac{\pi}{4} \end{cases} \tag{47}$$

where atan2(y, x) is the four-quadrant inverse tangent function for point (x, y).

To relate these results from Equation 47 to the amplitude and phase channels obtained by the holographic reconstruction method, a background normalization procedure is required, where the background region is normalized to have a unit amplitude and zero phase. For the background region, the output field can be written as:

$$\begin{aligned} \mathbf{E}_{\text{bg}} &= \begin{bmatrix} \cos^2\beta & \cos\beta\sin\beta \\ \cos\beta\sin\beta & \sin^2\beta \end{bmatrix} e^{-\frac{i\pi}{4}} \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -i \end{bmatrix} \\ &= \frac{1}{\sqrt{2}}\left(\cos\beta - \sin\beta\right) e^{\frac{i\pi}{4}} \cdot \mathbf{k} \end{aligned} \tag{48}$$

Thus, we have:

$$\begin{cases} A_{\mathrm{bg}} = \dfrac{\sqrt{2}}{2}\left|\cos\beta - \sin\beta\right| \\ \varphi_{\mathrm{bg}} = \operatorname{atan2}\left[0, \cos\beta - \sin\beta\right] + \dfrac{\pi}{4} \end{cases} \tag{49}$$

Hence, the reconstructed amplitude and phase can be expressed as:

$$\begin{cases} A_{\mathrm{recon}} = \dfrac{A_{\mathrm{out}}}{A_{\mathrm{bg}}} = \dfrac{\sqrt{a^2+b^2}}{\left|\cos\beta - \sin\beta\right|} \\ \varphi_{\mathrm{recon}} = \varphi_{\mathrm{out}} - \varphi_{\mathrm{bg}} = \operatorname{atan2}\left(b, a\right) - \dfrac{\eta}{2} - \dfrac{\pi}{2} - \operatorname{atan2}\left[0, \cos\beta - \sin\beta\right] \end{cases} \tag{50}$$

Error analysis of DL-HPM

The first direct source of error comes from the image co-registration. Any co-registration error negatively affects both the quantitative analysis as well as the training of the network. As was demonstrated in Figure 1, we used two image modalities to capture the image training and testing dataset. An accurate sub-pixel-level elastic co-registration algorithm is required to align the images from the two-imaging system. However, as the MSU crystals are much more rare than blood cells, the registration algorithm may focus only on aligning the cells at some certain regions and leaving a relatively large registration error for MSU crystals. Thus, Fig. 4.4 was plotted with another step of local registration to only align the individual crystal (a cross section without local registration is shown in Supporting Information S4.1). In addition, when the images used to train the network are inaccurately co-registered, the ground truth does not match the network input, which causes the loss function to be applied incorrectly. This in general can reduce the performance of the network.

Another cause of error is revealed by examining larger crystals, as the larger crystals are more likely to be formed by multiple MSU crystals which are stacked or clustered next to one another. This can lead to a more complex relationship between the amplitude/phase channel to the retardance/orientation channel, as the input phase channel may start to present wrapping effects and Equation 15 is insufficient to characterize the final polarization state. This is compounded by the fact that larger objects are the minority in the training dataset, the error in these regions becomes higher as is shown in the orientation channel of the example from the 20-50 μm length category in Fig. 4.4.

The experimental procedure can also cause errors. Throughout the image acquisition procedure used to generate both the training and testing dataset, each sample slide had to be aligned when moving it between two different imaging systems (each sample needs to be roughly aligned at the same angle relative to the polarizer-analyzer pair in each imaging system). Fig. S4.2 shows an example of what happens when the samples used by both the training and testing dataset are randomly aligned. In this figure, the reconstructed orientation channel performance is clearly degraded as the pseudo color does not matching the target. In addition to changes in the imaging procedure, if a change is made to the sample preparation procedure, the network may not generalize to the new sample. It is important that the network is trained using samples which are similar to those it is being tested on, as was shown in the results section.

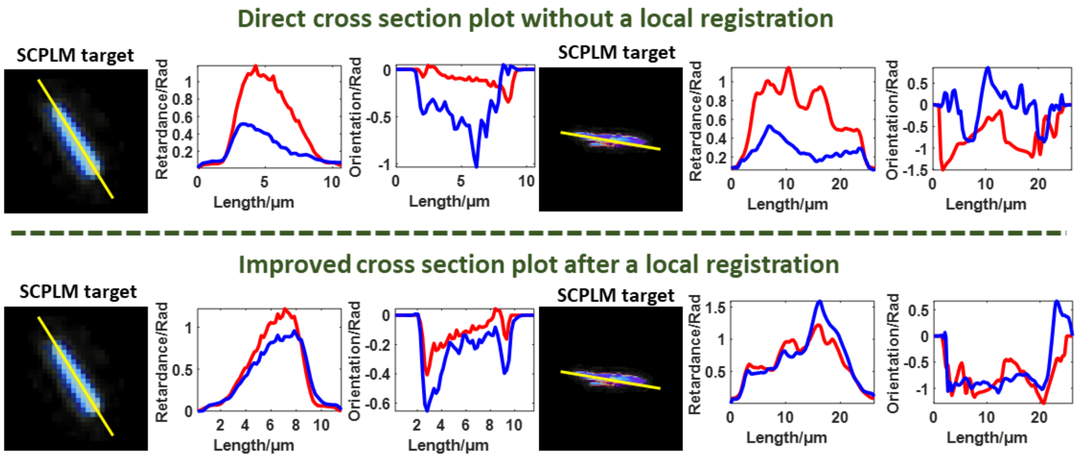


Figure S4.1 Comparison of cross section plots of retardance and orientation between DL-HPM output and SCPLM target with and without local registration. The blue curve represents the output for DL-HPM, while the red curve represents the target from SCPLM.

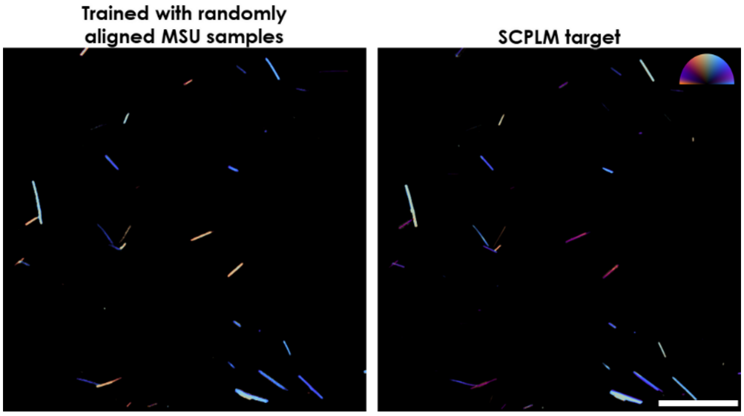
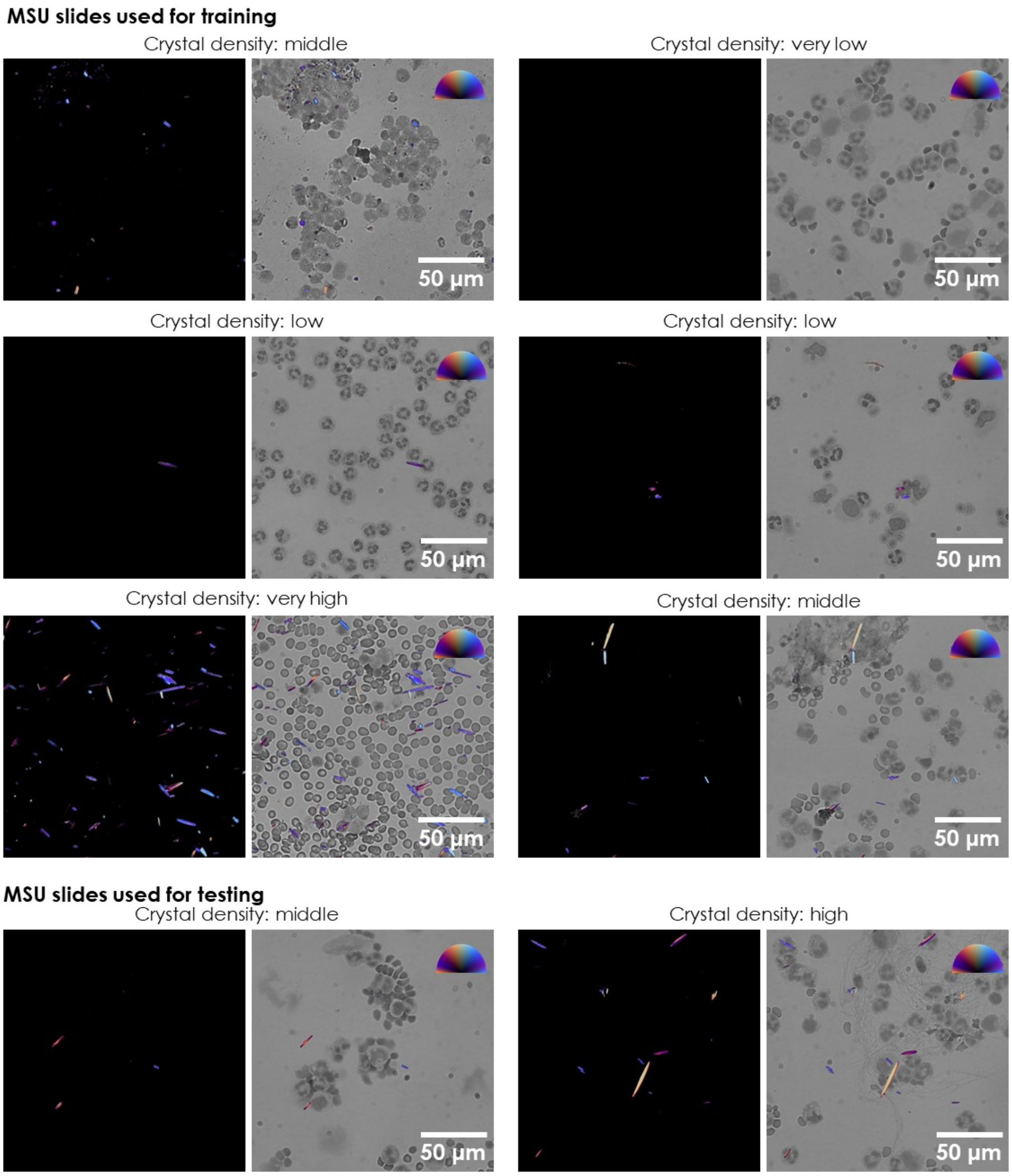


Figure S4.2 Result of DL-HPM when the training samples are not aligned. Color bar: from left to top to right, respectively, represents π, π/2, 0 Rad in the orientation channel. Retardance is represented by the distance from the center of the color bar, ranging from 0 to 0.4. Scale bar: 50 µm.

MSU dataset

The MSU dataset used for training and testing of DL-HPM contains 8 MSU slides obtained from different de-identified patients with different sample conditions in terms of the crystal density and dominant cell types. Figure S4.1 shows an example field of view for each of these slides obtained using the SCPLM method.



MSU slides used for training

Crystal density: middle

Crystal density: very low

Crystal density: low

Crystal density: low

Crystal density: very high

Crystal density: middle

MSU slides used for testing

Crystal density: middle

Crystal density: high

Figure S4.3 Example fields of view of each MSU slide used for training and testing of DL-HPM. Color bar: from left to top to right, respectively, represents π, π/2, 0 Rad in the orientation channel. Retardance is represented by the distance from the center of the color bar, ranging from 0 to 0.4.

Chapter 5 . Digital staining of label-free tissue using quantitative phase imaging

5.1 Introduction

Quantitative phase imaging (QPI) is a rapidly developing field, with a history of several decades in development [1,118]. QPI is a label-free imaging technique, which generates a quantitative image of the optical-path-delay through the specimen. Other than being label-free, QPI utilizes low-intensity illumination, while still allowing a rapid imaging time, which reduces phototoxicity in comparison to e.g., commonly-used fluorescence imaging modalities. QPI can be performed on multiple platforms and devices [5,9,18,119,120], from ultra-portable instruments all the way to custom-engineered systems integrated with standard microscopes, with different methods of extracting the quantitative phase information. QPI has also been recently used for the investigation of label-free thin tissue sections [1,121], which can be considered as a weakly scattering phase object, having limited amplitude contrast modulation under brightfield illumination.

Although QPI techniques result in quantitative contrast maps of label-free objects, the current clinical and research gold standard is still mostly based on brightfield imaging of histochemically labeled samples. The staining process dyes the specimen with colorimetric markers, revealing cellular and sub-cellular morphological information of the sample under brightfield microscopy. As an alternative, QPI has been demonstrated for the inference of local scattering coefficients of tissue samples [121,122]; for this information to be adopted as a diagnostic tool, some of the obstacles include the requirement of retraining experts and competing with a growing number of machine learning-based image analysis software [123,124], which utilize vast amounts of stained tissue images to perform e.g., automated diagnosis, image segmentation, or classification, among other tasks. One possible way to bridge the gap between QPI and standard imagebased diagnostic modalities is to perform digital (i.e., virtual) staining of phase images of label-free samples to match the images of histochemically-stained samples. One previously used method for digital staining of tissue sections involves the acquisition of multi-modal, nonlinear microscopy images of the samples, while applying staining regents as part of the sample preparation, followed by a linear approximation of the absorption process to produce a pseudo-Hematoxylin and Eosin (H&E) image of the tissue section under investigation [125–127].

As an alternative to model-based approximations, deep learning has recently been successful in various computational tasks based on a data-driven approach, solving inverse problems in optics, such as superresolution [26,27,30], holographic image reconstruction and phase recovery [14,76,128,129], tomography [130], Fourier ptychographic microscopy [131], localization microscopy [28,29,132] and ultra-short pulse reconstruction [133], among others. Recently, the application of deep learning for virtual staining of autofluorescence images of nonstained tissue samples has also been demonstrated [66]. Following on the success of these previous results, here we demonstrate that deep learning can be used for digital staining of label-free thin tissue sections using their quantitative phase images. For this image transformation between the phase image of a labelfree sample and its stained brightfield image, which we term as PhaseStain, we used a deep neural network trained using the Generative Adversarial Network (GAN) framework [72]. Conceptually, PhaseStain (see Fig. 5.1) provides an image that is the digital equivalent of a brightfield image of the same sample after the chemical staining process; stated differently it transforms the phase image of a weakly scattering object (e.g., a label-free thin tissue section, which exhibits low amplitude modulation under visible light) into an amplitude object information, presenting the same color features that are observed under a brightfield microscope, after the chemical staining process.

We experimentally demonstrated the success of our PhaseStain approach using label-free sections of human skin, kidney and liver tissue that were imaged by a holographic microscope, matching the brightfield microscopy images of the same tissue sections stained with H&E, Jones' stain, and Masson's trichrome stain, respectively.

The deep learning-based virtual-staining of label-free tissue samples using quantitative phase images provide another important example of the unique opportunities enabled by data-driven image transformations. We believe that the PhaseStain framework will be instrumental for QPI community to further strengthen the uses of label-free QPI techniques [134–138] for clinical applications and biomedical research, helping to eliminate the need for chemical staining, reduce sample preparation associated time, labor and related costs.

Part of this chapter has been published in:

Rivenson, Y., Liu, T., Wei, Z., Zhang, Y., de Haan, K., & Ozcan, A. (2019). PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. Light: Science & Applications, 8(1), 1-11.

5.2 Results

We trained 3 deep neural network models, which correspond to the 3 different combinations of tissue and stain types, i.e., H&E for skin tissue, Jones' stain for kidney tissue and Masson's trichrome for liver tissue. Following the training phase, these 3 trained deep networks were blindly tested on holographically reconstructed quantitative phase images (see the Methods section) that were not part of the network's training set. Figure 2 shows our results for virtual H&E staining of a phase image of a label-free skin tissue section, which confirms discohesive tumor cells lining papillary structures with dense fibrous cores. Additional results for virtual staining of quantitative phase images of label-free tissue sections are illustrated in Fig. 5.3, for kidney (digital Jones' staining) and liver (digital Masson's Trichrome staining). These virtually stained quantitative phase images show sheets of clear tumor cells arranged in small nests with a delicate capillary bed for the kidney tissue section, and a virtual trichrome stain highlighting normal liver architecture without significant fibrosis or inflammation, for the liver tissue section.

These deep learning-based virtual staining results presented in Figs. 5.2 and 5.3 visually demonstrate the high-fidelity performance of the GAN-based staining framework. To further shed light on this comparison between the PhaseStain results and the corresponding brightfield images of the chemically stained tissue samples, we quantified the structural similarity (SSIM) index of these two sets of images using:

$$\mathrm{SSIM}(U_1,U_2)=\frac{1}{3}\sum_{i=1,2,3}\frac{(2\mu_{1,i}\mu_{2,i}+2\sigma_{1,2,i}+c_2)}{(\mu_{1,i}^2+\mu_{2,i}^2+c_1)(\sigma_{1,i}^2+\sigma_{2,i}^2+c_2)} \tag{51}$$

where U_1, U_2 are the PhaseStain output and the corresponding brightfield reference image, respectively, $\mu_{k,i}$ and $\sigma_{k,i}$ are the mean and the standard deviation of each image U_k ($k = 1,2$), respectively, and index i refers to the RGB channels of the images. The cross-variance between the i-th image channels is denoted with $\sigma_{1,2,i}$ and c_1, c_2 are stabilization constants used to prevent division by a small denominator. The result of this analysis revealed that the SSIM was 0.8113, 0.8141 and 0.8905, for the virtual staining results corresponding to the skin, kidney and liver tissue samples, respectively, where the analysis was performed on ~10 Megapixel images, corresponding to a field-of-view (FOV) of ~1.47 mm^2 for each sample.

Next, to evaluate the sensitivity of the network output to phase noise in our measurements, we performed a numerical experiment on the quantitative phase image of a label-free skin tissue, where we added noise in the following manner:

$$\tilde{\phi}(m,n)=\phi(m,n)+\delta\phi(m,n)=\phi(m,n)+\beta r(m,n)*\frac{1}{2\pi L^2}\exp\left\{-(m^2+n^2)\Delta^2/[2(L\Delta)^2]\right\} \tag{52}$$

where $\tilde{\phi}$ is the resulting noisy phase distribution (i.e., the image under test), ϕ is the original phase image of the skin tissue sample, r is drawn from a normal distribution N(0,1) , β is the perturbation coefficient, L is the Gaussian filter size/width and Δ is the pixel size, which spatially smoothens the random noise into isotropic patches, as shown in Fig. 5.4. We choose these parameters such that the overall phase signal-to-noise-ratio (SNR) is statistically identical for all the cases and made sure that no phase wrapping occurs. We then used 10 random realizations of this noisy phase image for 4 combinations of (β, L) values to generate $\tilde{\phi}$ which was used as input to our trained deep neural network. The deep network inference fidelity for these noisy phase inputs is reported in Fig. 5.4, which reveals that it is indeed sensitive to local phase variations and related noise, and it improves its inference performance as we spatially extend the filter size, L (while the SNR remains fixed). In other words, the PhaseStain network output is more impacted by small scale variations, corresponding to e.g., the information encoded in the morphology of the edges or refractive index discontinuities (or sharp gradients) of the sample. We also found that for a kernel size of $L\Delta$~3 μm, the SSIM remains unchanged (~0.8), across a wide range of perturbation coefficients, β. This result implies that the network is less sensitive to sample preparation imperfections, such as height variations and wrinkles in the thin tissue section, which naturally occur during the preparation of the tissue section.

5.3 Discussion

The training process of a PhaseStain network needs to be performed only once, following which, the newly acquired quantitative phase images of various samples are blindly fed to the pre-trained deep network to output a digitally-stained image for each label-free sample, corresponding to the image of the same sample FOV, as it would have been imaged with a brightfield microscope, following the chemical staining process. In terms of the computation speed, the virtual staining using PhaseStain takes 0.617 sec on average, using a standard desktop computer equipped with a dual-GPU for a FOV of ~0.45 mm^2 , corresponding to ~3.22 Megapixels (see the implementation details in the Methods section). This fast inference time, even with relatively modest computers, means that the PhaseStain network can be easily integrated with a QPI-based whole slide scanner, since the network can output virtually-stained images in small patches while the tissue is still being scanned by an automated microscope, to simultaneously create label-free QPI and digitally-stained whole slide images of the samples.

The proposed technology has the potential to save time, labor and costs, by presenting an alternative to the standard histochemical staining workflow used in clinical pathology. As an example, one of the most common staining procedures (i.e., H&E stain) takes on average ~45 min and costs approximately $2-5, while the Masson's Trichrome staining procedure takes ~2-3 hours, with costs that range between $16-35, and often requires monitoring of the process by an expert, which is typically conducted by periodically examining the specimen under a microscope. In addition to saving time and costs, by circumventing the staining procedure, the tissue constituents would not be altered; this means the unlabeled tissue sections can be preserved for later analysis, such as matrix-assisted laser desorption ionization (MALDI) by micro-sectioning of specific areas [139] for molecular analysis or micro-marking of sub-regions which can be labeled with specific immunofluorescence tags or tested for personalized therapeutic strategies and drugs [140,141].

While in this study we trained 3 different neural network models to obtain optimal results for specific tissue and stain combinations, this does not pose a practical limitation for PhaseStain, since we can also train a more general digital staining model for a specific stain type (e.g. H&E, Jones' stain etc.) using multiple tissue types stained with it, at the cost of increasing the network size as well as the training and inference times [76]. Also, from clinical diagnostics perspective, the tissue type under investigation and the stain needed for its clinical examination are both known a priori, and therefore the selection of the correct neural network for each sample to be examined is straightforward to implement.

It is important to note that, in addition to the lensfree holographic microscope (see the Methods section) that we used in this work, PhaseStain framework can also be applied to virtually-stain the resulting images of various other QPI techniques, regardless of their imaging configuration, specific hardware or phase recovery method [1,18,33,120,142–144] that are employed.

One of the disadvantages of coherent imaging systems is “coherence-related image artifacts”, such as e.g., speckle noise, or dust or other particles creating holographic interference fringes, which do not appear in incoherent brightfield microscopy images of the sample samples. In Fig. 5.5, we demonstrate the image distortions that, for example, out-of-focus particles create on the PhaseStain output image. To reduce such distortions in the network output images, coherence-related image artifacts resulting from out-of-focus particles can be digitally removed by using a recently introduced deep learning-based hologram reconstruction method, which learns, through data, to attack or eliminate twin-image artifact as well as interference fringes resulting from out-of-focus or undesired objects [12,14].

While in this manuscript we demonstrated the applicability of PhaseStain approach to fixed paraffinembedded tissue specimen, our approach should be also applicable to frozen tissue sections, involving other tissue fixation methods as well (following a similar training process as detailed in the Methods section). Also, while our method was demonstrated for thin tissue sections, QPI has been shown to be valuable to image cells and smear samples (such as blood and Pap smears) [1,33], and PhaseStain technique would also be applicable to digitally stain these types of specimen.

To summarize, our presented results demonstrate some of the emerging opportunities created by deep learning for label-free quantitative phase imaging. The phase information resulting from various coherent imaging techniques can be used to generate a virtually stained image, translating the phase images of weakly scattering objects such as thin tissue sections into images that are equivalent to the brightfield images of the same samples, after the histochemical labeling. PhaseStain framework, in addition to saving time and cost associated with the labeling process, has the potential to further strengthen the use of label-free QPI techniques in clinical diagnostics workflow, while also preserving tissues for e.g., subsequent molecular and genetic analysis.

5.4 Materials and methods

Sample preparation and imaging

All the samples that were used in this study were obtained from the Translational Pathology Core Laboratory (TPCL) and were prepared by the Histology Lab at UCLA. They were obtained after deidentification of the patient related information and were prepared from existing specimen. Therefore, this work did not interfere with standard practices of care or sample collection procedures.

Following formalin-fixing paraffin-embedding (FFPE), the tissue block is sectioned using a microtome into ~2-4 μm thick sections. This step is only needed for the training phase, where the transformation from a phase image into a brightfield image needs to be statistically learned. These tissue sections are then deparaffinized using Xylene and mounted on a standard glass slide using CytosealTM (ThermoFisher Scientific, Waltham, MA USA), followed by sealing of the specimen with a coverslip. In the learning/training process, this sealing step presents several advantages: protecting the sample during the imaging and sample handling processes, also reducing artifacts such as e.g., sample thickness variations. Following the sample preparation, the specimen was imaged using an on-chip holographic microscope to generate a quantitative phase image (detailed in the next sub-section).

Following the QPI process, the label-free specimen slide was put into Xylene for ~48 hours, until the coverslip can be removed without introducing distortions to the tissue. Once the coverslip is removed the slide was dipped multiple times in absolute alcohol, 95% alcohol and then washed in D.I. water for ~1 min. Following this step, the tissue slides were stained with H&E (skin tissue), Jones' stain (kidney tissue) and Masson's trichrome (liver tissue) and then coverslipped. These tissue samples were then imaged using a brightfield automated slide scanner microscope (Aperio AT, Leica Biosystems) with a 20×/0.75NA objective (Plan Apo), equipped with a 2× magnification adapter, which results an effective pixel size of ~0.25 μm.

Quantitative phase imaging

Lensfree imaging setup: Quantitative phase images of label-free tissue samples were acquired using an in-line lens-free holography setup [33]. A light source (WhiteLase Micro, NKT Photonics) with a center wavelength at 550 nm and a spectral bandwidth of ~2.5nm was used as the illumination source. The uncollimated light emitted from a single-mode fiber was used for creating a quasi-plane-wave that illuminated the sample. The sample was placed between the light source and the CMOS image sensor chip (IMX 081, Sony, pixel size of 1.12 μm) with a source-to-sample distance (z1) of 5~10 cm and a sample-to-sensor distance (z2) of 1-2 mm. This on-chip lensfree holographic microscope has submicron resolution with an effective pixel size of 0.37 μm, covering a sample FOV of ~20 mm^2 (which accounts for the entire active area of the sensor). The positioning stage (MAX606, Thorlabs, Inc.), that held the CMOS sensor, enabled 3D translation of the imager chip for performing pixel super-resolution (PSR) [5,17,33] and multi-height based iterative phase recovery [10,33]. All imaging hardware was controlled automatically by LabVIEW.

Pixel super-resolution (PSR) technique: To synthesize a high-resolution hologram (with a pixel size of ~0.37 μm) using only the G1 channel of the Bayer pattern (R, G1, G2, and B), a shift-and-add based PSR algorithm was applied [17,145]. The translation stage that holds the image sensor was programmed to laterally shift on a 6×6 grid with sub-pixel spacing at each sample-to-sensor distance. A low-resolution hologram was recorded at each position and the lateral shifts were precisely estimated using a shift estimation algorithm [33]. This step results in 6 non-overlapping panels that were each padded to a size of 4096×4096 pixels, and were individually phase-recovered, which is detailed next.

Multi-height phase recovery: Lensfree in-line holograms at eight sample-to-sensor distances were captured. The axial scanning step size was chosen to be 15 µm. Accurate z-steps were obtained by applying a holographic autofocusing algorithm based on the edge sparsity criterion ("Tamura of the gradient", i.e., ToG) [38]. A zero-phase was assigned to the object intensity measurement as an initial phase guess, to start the iterations. An iterative multi-height phase recovery algorithm41 was then used by propagating the complex field back and forth between each height using the transfer function of freespace [34]. During this iterative process, the phase was kept unchanged at each axial plane, where the amplitude was updated by using the square-root of the object intensity measurement. One iteration was defined as propagating the hologram from the 8th height (farthest from the sensor chip) to the 1st height (nearest to the sensor) then back propagating the complex field to the 8th height. Typically, after 10-30 iterations the phase is retrieved. For the final step of the reconstruction, the complex wave defined by the converged amplitude and phase at a given hologram plane was propagated to the object plane [4], from which the phase component of the sample was extracted.

Data preprocessing and image registration

An important step in our training process is to perform an accurate image registration, between the two imaging modalities (QPI and brightfield), which involves both global matching and local alignment steps. Since the network aims to learn the transformation from a label-free phase retrieved image to a histochemically-stained brightfield image, it is crucial to accurately align the FOVs for each input and target image pair in the dataset. We perform this cross-modality alignment procedure in four steps; steps 1,2 and 4 are done in Matlab (The

MathWorks Inc., Natick, MA, USA) and step 3 involves TensorFlow.

The first step is to find a roughly matched FOV between QPI and the corresponding brightfield image. This is done by first bicubic down-sampling the whole slide image (WSI) (~60k by 60k pixels) to match the pixel size of the phase retrieved image. Then, each 4096×4096-pixel phase image was cropped by 256 on each side (resulting in an image with 3584×3584 pixels) to remove the padding that is used for the image reconstruction process. Following this step, both the brightfield and the corresponding phase images are edge extracted using the Canny method [146], which uses double threshold to detect strong and weak edges on the gradient of the image. Then, a correlation score matrix is calculated by correlating each 3584x3584-pixel patch of the resulting edge image to the same size as the image extracted from the brightfield edge image. The image with the highest correlation score indicates a match between the two images, and the corresponding brightfield image is cropped out from the WSI. Following this initial matching procedure, the quantitative phase image and the brightfield microscope images are coarsely matched.

The second step is used to correct for potential rotations between these coarsely matched image pairs, which might be caused by a slight mismatch in the sample placement during the two image acquisition experiments (which are performed on different imaging systems, holographic vs. brightfield). This intensity-based registration step correlates the spatial patterns between the two images; phase image that is converted to unsigned integer format and the luminance component of the brightfield image were used for this multimodal registration framework implemented in Matlab. The result of this digital procedure is an affine transformation matrix, which is applied to the brightfield microscope image patch, to match it with the quantitative phase image of the same sample. Following this registration step, the phase and the corresponding brightfield images are globally aligned. A further crop of 64 pixels on each side to the aligned image pairs is used to accommodate for a possible rotation angle correction.

The third step involves the training of a separate neural network that roughly learns the transformation from quantitative phase images into stained brightfield images, which can help the distortion correction between the two image modalities in the fourth/final step. This neural network has the same structure as the network that was used for the final training process (see the next sub-section on GAN architecture and its training) with the input and target images obtained from the second registration step discussed earlier. Since the image pairs are not well aligned yet, the training is stopped early at only ~2000 iterations to avoid a structural change at the output to be learnt by the network. The output and target images of the network are then used as the registration pairs in the fourth step, which is an elastic image registration algorithm, used to correct for local feature registration [27].

GAN architecture and training

The GAN architecture that we used for PhaseStain is detailed in Table 5.1. Following the registration of the label-free quantitative phase images to the brightfield images of the histochemically stained tissue sections, these accurately aligned fields-of-view were partitioned to overlapping patches of 256×256 pixels, which were then used to train the GAN model. The GAN is composed of two deep neural networks, a generator and a discriminator. The discriminator network's loss function is given by:

$$l_{\text{discriminator}} = D(G(x_{\text{input}}))^2 + (1 - D(z_{\text{label}}))^2 \tag{53}$$

where *D*(.) and *G*(.) refer to the discriminator and generator network operator, input x denotes the input to the generator, which is the label-free quantitative phase image, and z_{label} denotes the brightfield image of the chemically stained tissue. The generator network, *G*, tries to generate an output image with the same statistical features as label z_{label}, while the discriminator, *D*, attempts to distinguish between the target and the generator output images. The ideal outcome (or state of equilibrium) will be when the generator's output and target images share an identical statistical distribution, where in this case, $D(G(x_{\text{input}}))$ should converge to 0.5. For the generator deep network, we defined the loss function as:

$$l_{\text{generator}} = L_1\{z_{\text{label}}, G(x_{\text{input}})\} + \lambda \times \text{TV}\{G(x_{\text{input}})\} + \alpha \times (1 - D(G(x_{\text{input}})))^2 \tag{54}$$

where $L1\{.\}$ term refers to the absolute pixel by pixel difference between the generator output image and its target, $\text{TV}\{.\}$ stands for the total variation regularization that is being applied to the generator output, and the last term reflects a penalty related to the discriminator network prediction of the generator output.

The regularization parameters (λ, α) were set to 0.02 and 2000 so that the total variation loss term, $\lambda \times \text{TV}\ \{G\ (x_{\text{input}})\ \}$, was ~2% of the L1 loss term, and the discriminator loss term, input 2 $\alpha \times\ (1\text{-}D\ (G(x_{\text{input}})))$ was ~98% of the total generator loss, generator $l_{\text{generator}}$.

For the generator deep neural network, we adapted the U-net architecture [147], which consists of a downsampling and an up-sampling path, with each path containing 4 blocks forming 4 distinct levels (see Table 5.1). In the down-sampling path, each residual block consists of 3 convolutional layers and 3 leaky rectified linear (*L*ReLU) units used as an activation function, which is defined as:

$$L\text{ReLU}(x) = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \tag{55}$$

At the output of each block, the number of channels is 2-fold increased (except for the first block that increases from 1 input channel to 64 channels). Blocks are connected by an average-pooling layer of stride 2 that down-samples the output of the previous block by a factor of 2 for both horizontal and vertical dimensions (as shown in Table 5.1).

In the up-sampling path, each block also consists of 3 convolutional layers and 3 *L*ReLU activation functions, which decrease the number of channels at its output by 4-fold. Blocks are connected by a bilinear up-sampling layer that up-samples the size of the output from the previous block by a factor of 2 for both lateral dimensions. A concatenation function with the corresponding feature map from the downsampling path of the same level is used to increase the number of channels from the output of the previous block by 2. The two paths are connected in the first level of the network by a convolutional layer which maintains the number of the feature maps from the output of the last residual block in the downsampling path (see Table 5.1). The last layer is a convolutional layer that maps the output of the upsampling path into 3 channels of the YCbCr color map.

The discriminator network consists of one convolutional layer, 5 discriminator blocks, an average pooling layer and two fully connected layers. The first convolutional layer receives 3 channels (YCbCr color map) from either the generator output or the target, and increases the number of channels to 64. The discriminator blocks consist of 2 convolutional layers with the first layer maintaining the size of the feature map and the number of channels, while the second layer increases the number of channels by 2- fold and decreases the size of the feature map by 4-fold. The average pooling layer has a filter size of 8×8, which results in a matrix with a size of (B, 2048), where B refers to the batch size. The output of this average pooling layer is then fed into two fully connected layers with the first layer maintaining the size of the feature map, while the second layer decreases the output channel to 1, resulting in an output size of (B, 1). The output of this fully connected layer is going through a sigmoid function indicating the probability that the 3-channel discriminator input is drawn from a chemically stained brightfield image. For the discriminator network, all the convolutional layers and fully connected layers are connected by *L*ReLU nonlinear activation functions.

Throughout the training, the convolution filter size was set to be 3×3. For the patch generation, we applied data augmentation by using 50% patch overlap for the liver and skin tissue images, and 25% patch overlap for the kidney tissue images (see Table 5.2). The learnable parameters including filters, weights and biases in the convolutional layers and fully connected layers are updated using an adaptive moment estimation (Adam) optimizer with learning rate 1×10-4 for the generator network and 1×10-5 for the discriminator network.

For each iteration of the discriminator, there were v iterations of the generator network; for liver and skin tissue training, v = max(5, floor(7-w/2)) where we increased w by 1 for every 500 iterations (w was initialized as 0). For the kidney tissue training, we used v = max(4, floor(6-w/2)) where we increased w by 1 for every 400 iteration. This helped us to train the discriminator not to overfit to the target brightfield images. We used a batch size of 10 for the training of liver and skin tissue sections, and 5 for the kidney tissue sections. The network's training stopped when the validation set's L1-loss did not decrease after 4000 iterations. A typical convergence plot of our training is shown in Fig. 5.6.

Implementation details

The number of image patches that were used for training, the number of epochs and the training schedules are shown in Table 5.2. The network was implemented using Python version 3.5.0, with TensorFlow framework version 1.7.0. We implemented the software on a desktop computer with a Core i7-7700K CPU @ 4.2GHz (Intel) and 64GB of RAM, running a Windows 10 operating system (Microsoft). Following the training for each tissue section, the corresponding network was tested with 4 image patches of 1792×1792 pixels with an overlap of ~7%. The outputs of the network were then stitched to form the final network output image of 3456×3456 pixels (FOV ~1.7 mm^2), as shown in e.g., Fig. 5.2. The network training and testing were performed using dual GeForce GTX 1080Ti GPUs (NVidia).

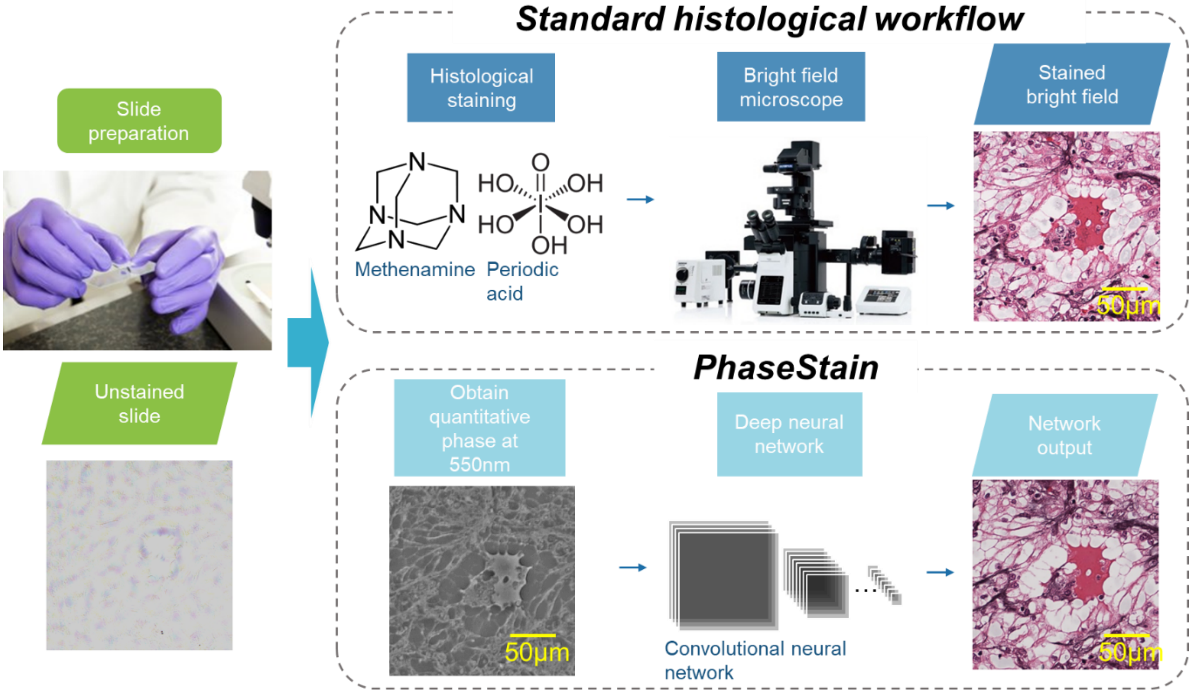


Figure 5.1 PhaseStain workflow. Quantitative phase image of a label-free specimen is virtually stained by a deep neural network, bypassing the standard histochemical staining procedure that is used as part of clinical pathology.

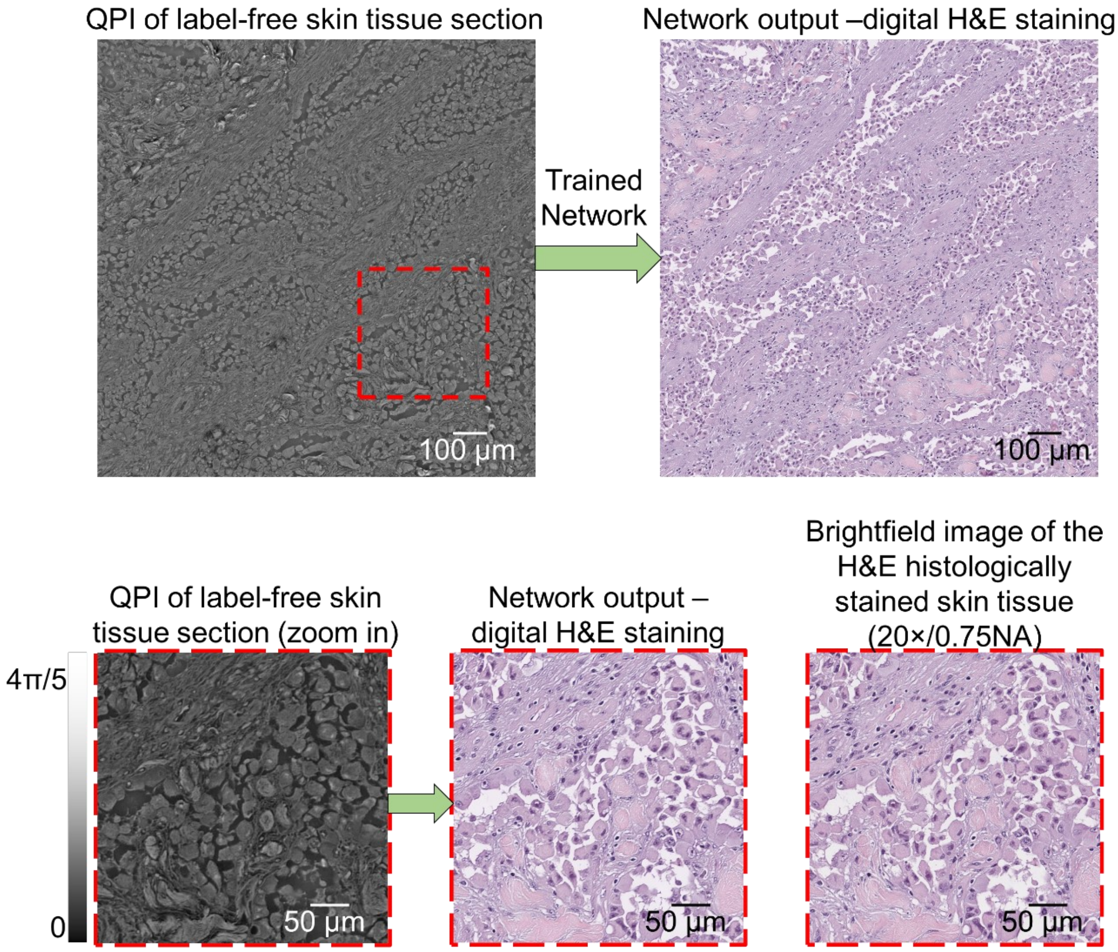


Figure 5.2 Virtual H&E staining of label-free skin tissue using PhaseStain framework.

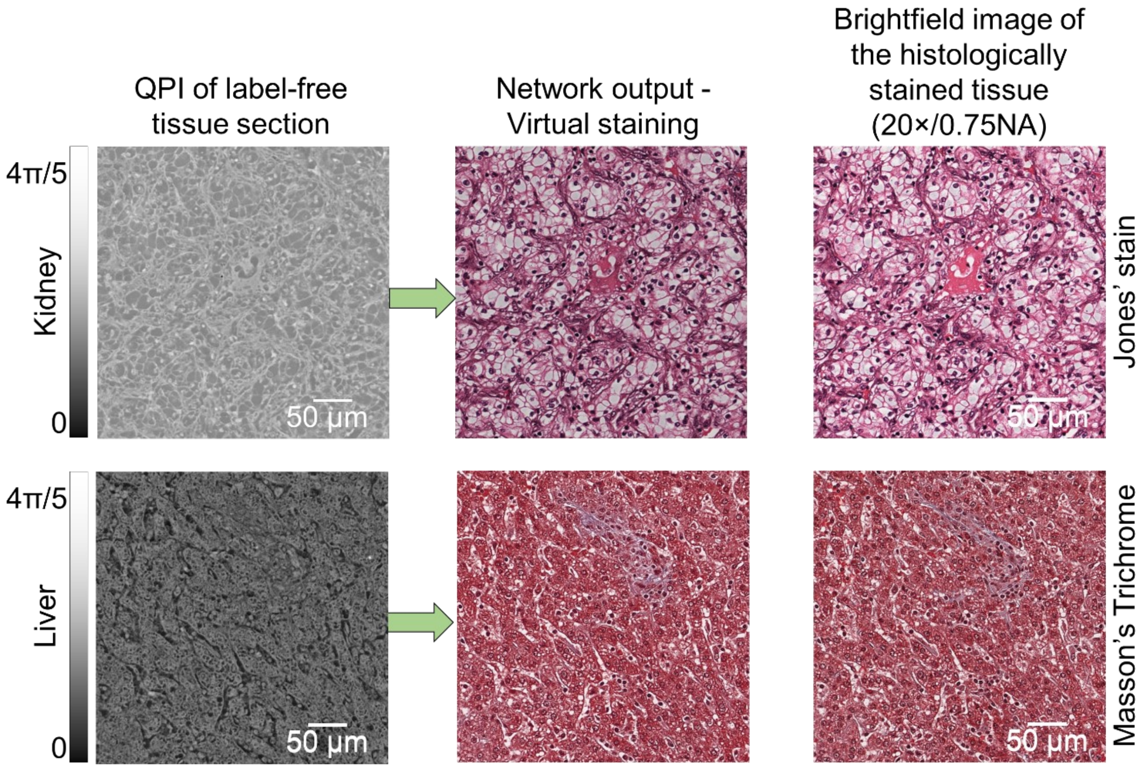


Figure 5.3 PhaseStain based virtual staining of label-free kidney tissue (Jones' stain) and liver tissue (Masson's Trichrome).

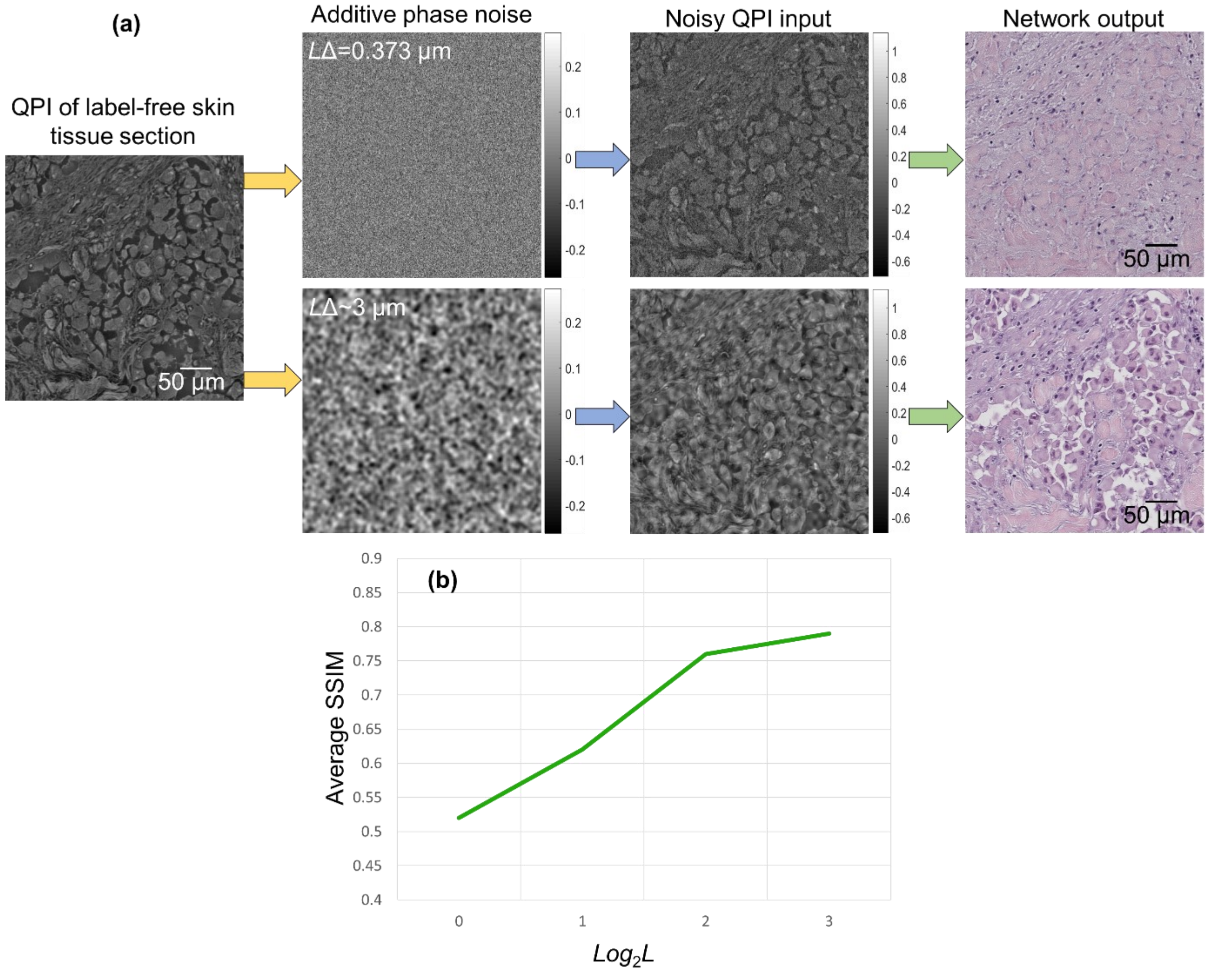


Figure 5.4 (a) PhaseStain results for noisy phase input images (ground truth shown in Fig. 5.2). Top row: LΔ~0.373 μm; second row: LΔ~3 μm. **(b) Analysis of the impact of phase noise on the inference quality of PhaseStain (quantified using SSIM), as a function of the Gaussian filter length, L (see Eq. 52).**

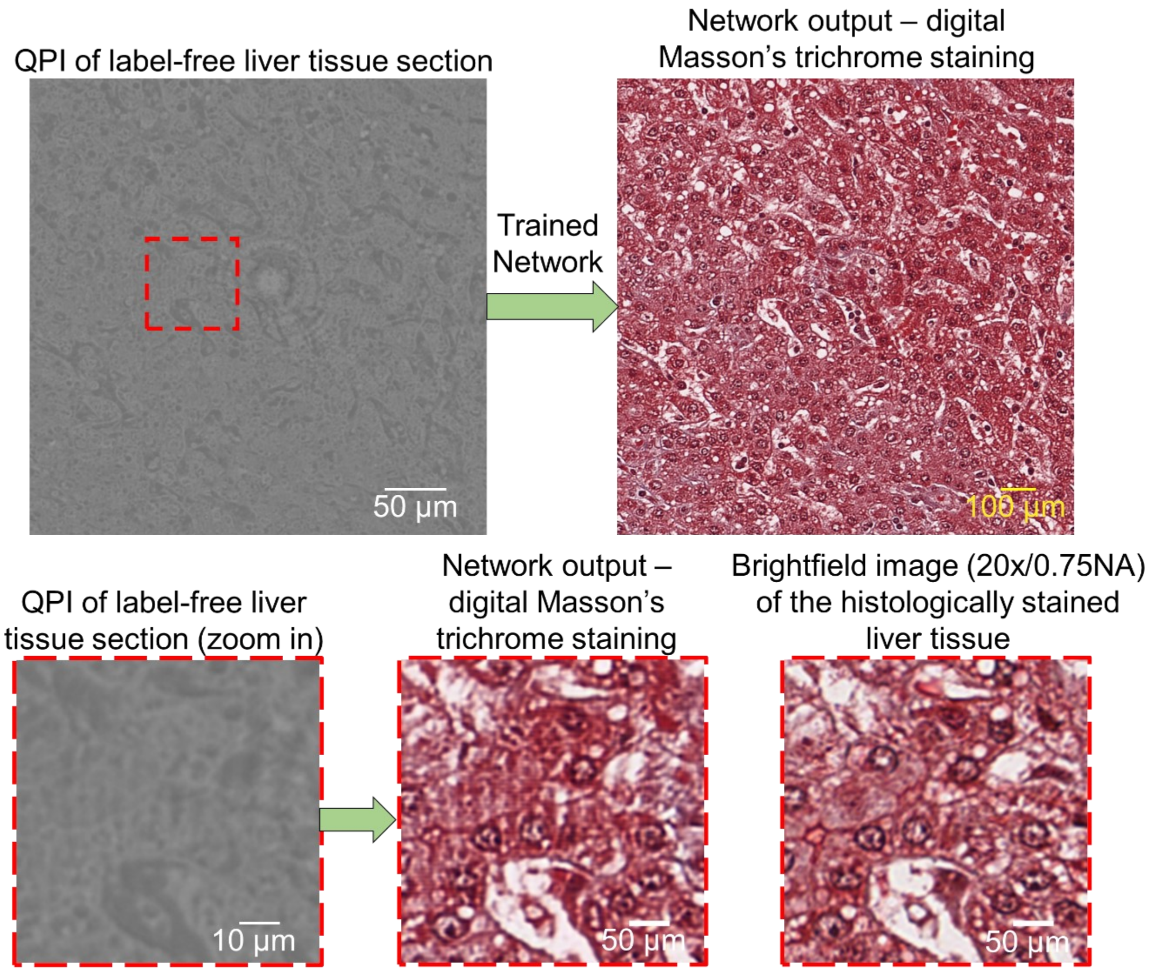


Figure 5.5 The impact of holographic fringes resulting from out-of-focus particles on the deep neural network's digital staining performance.

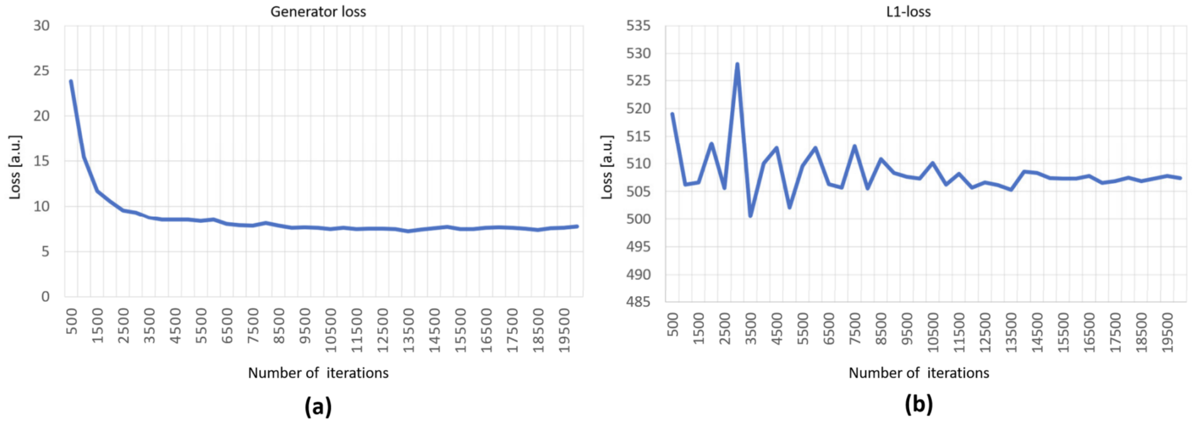


Figure 5.6 PhaseStain convergence plots for the validation set of the digital H&E staining of the skin tissue. (a) L1-loss with respect to the number of iterations. (b) Generator loss, generator ℓ with respect to the number of iterations.

Chapter 6. Rare cell detection using holographic time-lapse imaging

6.1 Introduction

Rare cell detection aims to identify a sufficient number of low-abundance cells within a vast majority of background cells, which typically requires the processing of large volumes of biological sample. The detection and enumeration of these rare cells are vital for disease diagnostics, the evaluation of disease progression, and the characterization of immune response. [148–150] For instance, circulating foetal cells present in maternal blood are recognized as a source of foetal genomic DNA, and their isolation is crucial for the implementation of routine prenatal diagnostic testing. [151] As another example, antigen-specific T cells in peripheral blood play a central role in mediating immune response and the formation of immunological memory, which could lead to the prediction of immune protection and diagnosis of immune-related diseases. [152] Circulating endothelial cells with a mature phenotype are increased in patients with certain types of cancer and several pathological conditions, indicating their potential as disease markers. [153] Circulating tumour cells (CTCs) are implicated in various stages of cancer and have therefore been collected to study their role in the metastatic cascade and to predict patient outcomes from both the disease and treatments received. [154] To highlight yet another example, haematopoietic stem and progenitor cells, which reside predominantly in bone marrow with low numbers, also found in peripheral blood, possess the unique capacity for self-renewal and multilineage differentiation, and their trafficking in blood may be connected to disease processes. [155]

The specific and sensitive detection of these rare cells in human blood and other bodily fluids is therefore of great interest. However, millions of events need to be acquired to obtain a sufficient number of these low-abundance cells (e.g., typically <1000 target cells per millilitre of blood [156]). The direct detection of rare cells from whole blood requires the processing of large amounts of patient sample (e.g., up to a few hundred millilitres [157]), which is both unrealistic and time consuming. To alleviate this issue, highly specific labelling methods are often used before detection for sample purification/enrichment to facilitate rapid detection and processing. [152,157] Among these labelling techniques, the use of colloidal magnetic particles as labelling reagents offers benefits in forming stable suspensions, fast reaction kinetics, [157] and minimum damage to the target cells, with high viability retained. [158]

Motivated by these important needs and the associated challenges, various technologies have been developed and employed for detecting rare cells in blood. Most of these existing detection methods involve three steps: capture, enrichment and detection. [159] The capture and enrichment steps use a number of methods, such as barcoded particles, [160] magnetic beads, [161] micro-machines, [162] microfluidic chips, [163] and density gradient centrifugation. [159,164] Following the enrichment step, these rare cells can be detected via commonly used techniques such as immunofluorescence, [165,166] electrical impedance, [167] and Raman scattering [168] measurements, among others. Notably, commercial products for rare cell detection, such as the CellSearch system, [169] which automates magnetic labelling, isolation, fluorescence labelling, and automated counting, are generally high-cost, limiting their adoption worldwide. [159] Therefore, cost-effective, reliable and high-throughput rare cell detection techniques are urgently needed to improve the early diagnosis of diseases, including cancer, so that earlier treatments can be carried out, helping us to improve patient outcomes while also reducing healthcare costs. [170,171]

The recent advances in machine learning and, specifically, deep learning have pushed the frontiers of biomedical imaging and image analysis, [128,172–184] enabling rapid and accurate pathogen detection [185–188] and computer-assisted diagnostic methods. [189–193] Powered by deep learning, we demonstrate here that speckle imaging using lensless chip-scale microscopy can be employed for the specific and sensitive detection of rare cells in blood with low cost and high throughput. This novel cell detection and cytometry technique is based on magnetically modulated lensless speckle imaging, which specifically labels rare cells of interest using magnetic particles attached to surface markers of interest and generates periodic and well-controlled motion on target cells by alternating the external magnetic field applied to a large sample volume. The holographic diffraction and the resulting speckle patterns of the moving cells are then captured using a compact and cost-effective on-chip lensless imager (Fig. 6.1) and are computationally analysed by a deep-learning-based algorithm to rapidly detect and accurately identify the rare cells of interest in a high-throughput manner based on their unique spatio-temporal features. Although previous work has employed the idea of using magnetic modulation for enhancing fluorescence detection, [194,195] our work is the first of its kind for combining magnetic modulation, lensless imaging and deep learning to create a unique cytometer that does not require additional labelling (e.g., fluorescence) or custom-designed molecular probes.

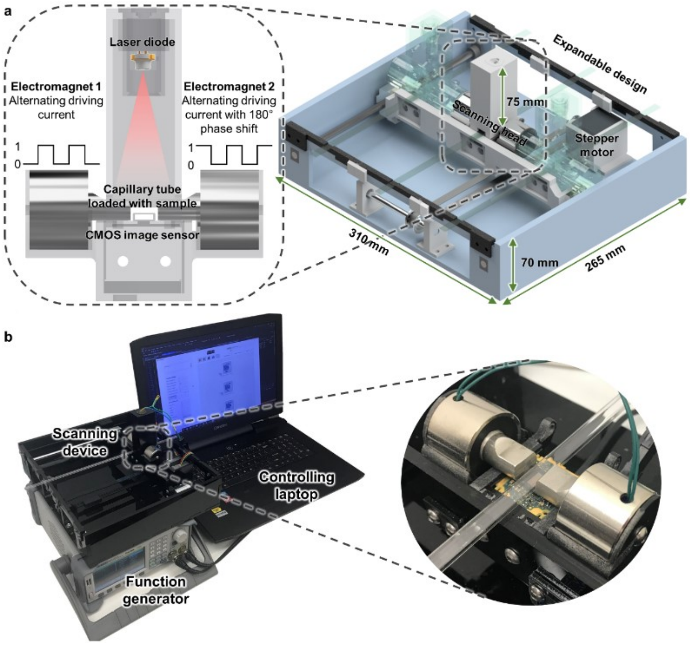


Figure 6.1 Schematics and photos of the computational cytometer. (a) A magnetically modulated lensless imaging module (inset) that includes a lensless holographic microscope and two electromagnets driven by

alternating currents with opposite phase. The fluid sample that contains magnetic-bead-conjugated cells of interest is loaded into a capillary tube. The imaging module is mounted to a linear motion stage to scan along the sample tube to record holographic images of each section of the tube. (b) A laptop computer is used to control the device and acquire data. A function generator and a power supply, together with custom-designed circuitry, are used to provide the power and driving current for the linear motion stage and electromagnets.

As shown in Fig. 6.1, we built a portable prototype of this computational cytometer for rare cell detection. Our magnetically modulated speckle imaging module includes a lensless in-line holographic microscope [51,52,55,60,187,196–199] and two oppositely positioned electromagnets (Fig. 6.1a-b inset). The lensless microscope contains a laser diode (650 nm wavelength) to illuminate the sample from ~5-10 cm above, and a complementary metal–oxide–semiconductor (CMOS) image sensor is placed ~1 mm below the sample for acquisition of a high-frame-rate video to monitor the spatio-temporal evolution of the sample containing the target cells of interest. Because the light-source-to-sample distance is much greater than the sample-to-image-sensor distance, the optical design has a unit magnification, and the field of view (FOV) of a single image is equal to the active area of the image sensor (which can be 10-30 mm^2 using the standard CMOS imagers employed in digital cameras and mobile phones). To increase the screening throughput, target cells are enriched using magnetic separation and loaded inside a capillary tube for imaging (Figs. 6.1-6.2). Magnetic enrichment alone leads to a background of unlabelled cells, bead clusters, or weakly labelled cells that are also captured, such that further discrimination of the target cells within this background information is needed to accurately identify and count the rare cells. The imaging module is mounted onto a custom-made linear translation stage and is translated along the direction of the sample tube to capture a holographic video for each section of the sample tube. During the imaging at each section, the electromagnets are supplied with alternating current with a 180° phase difference to exert an alternating pulling force to the magnetic-bead-conjugated cells in the sample, which causes them to oscillate at the same frequency as the driving current. Extension rods made of permalloy were designed and utilized to enhance the magnetic force at the sample location by ~40-fold (see the Methods section and Fig. S6.1). The holographic diffraction patterns that are cast by the magnetically modulated target cells are captured using the image sensor and are transferred to a laptop computer. A computational motion analysis (CMA) algorithm [187] and a densely connected pseudo-3D convolutional neural network structure (P3D CNN) [200] then analyse the holographic image sequence that contains the 3D dynamic information from the oscillating cells, which allows rapid and specific detection of the target cells.

The current prototype (Fig. 6.1) screens ~0.942 mL of fluid sample, corresponding to ~1.177 mL of whole blood sample before enrichment, in ~7 min (Fig. 6.2), while costing only ~$750 for the raw materials (excluding the function generator, power supply and laptop computer) and weighing ~2.1 kg. The platform with a single imaging channel can be expanded to parallel imaging channels by mounting several imaging modules onto the same linear stage, as shown in Fig. 6.1a (semi-translucent illustrations).

Part of this chapter has been published in:

Zhang, Y., Ouyang, M., Ray, A., Liu, T., Kong, J., Bai, B., ... & Tsai, K. (2019). Computational cytometer based on magnetically modulated coherent imaging and deep learning. Light: Science & Applications, 8(1), 1-15.

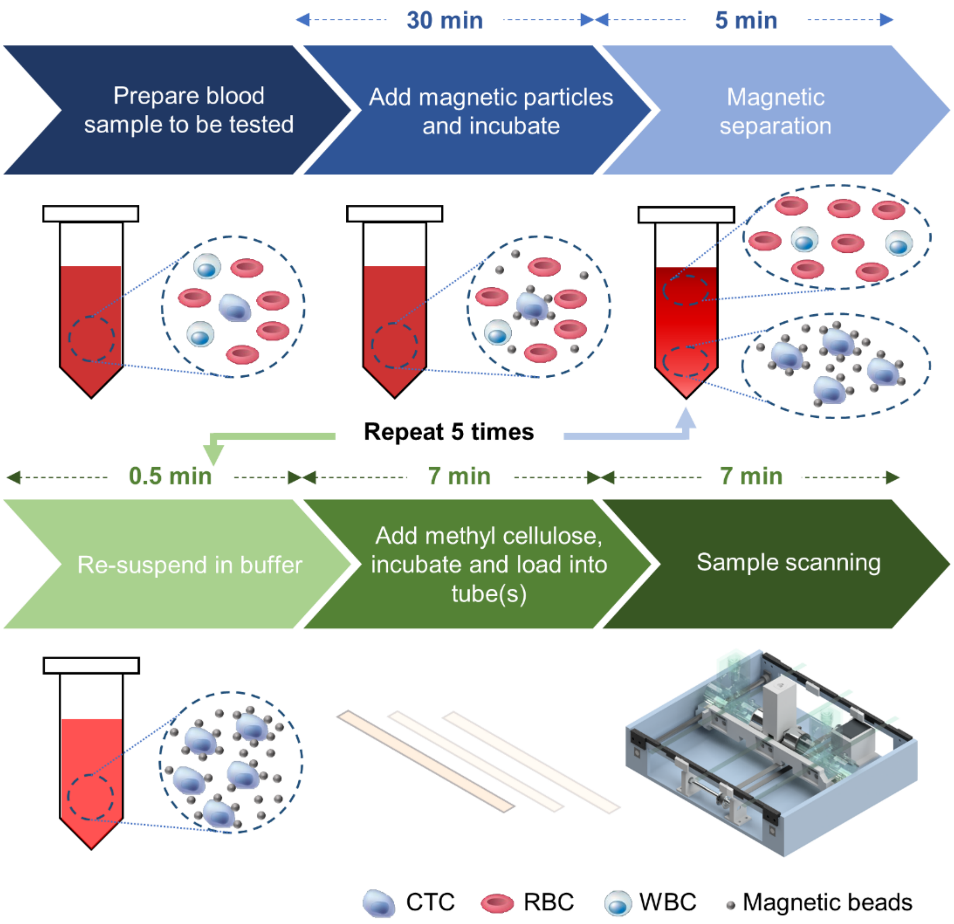


Figure 6.2 Sample preparation and imaging procedures. The sample preparation time before scanning is approximately 1 hour, with the first 30 min dedicated to passive incubation, which does not require supervision.

The performance of our technique was tested by detecting a model rare cell system of spiked MCF7 cancer cells in human blood. We demonstrate that our technique has a limit of detection (LoD) of 10 cells per mL of whole blood using only a single imaging channel. Because the current LoD is mainly limited by the screening volume, we expect that the LoD can be further improved by including additional parallel imaging channels and increasing the sample volume that is screened.

6.2 Results

Characterization of the oscillation of bead-cell conjugates under alternating magnetic force

Our detection technique capitalizes on the periodic oscillatory motion of the target cells of interest, with a large number of labelling magnetic particles, to specifically detect them with high throughput. We designed a magnetic actuator to exert periodic and alternating magnetic force on the magnetic particles bound to these cells of interest (Fig. 6.1). To exert sufficient magnetic force on each labelled cell, we designed and machined extension rods that were made with magnetically soft permalloy, which were attached to the electromagnets to enhance the magnetic force at the sample location by ~40-fold with minimal magnetic hysteresis (see the Methods section and Fig. S6.1).

The movement of MCF7 cells conjugated with Dynabeads was recorded by mounting the magnetic actuator and the labelled cells onto a 40× 0.6NA benchtop microscope (see Fig. 6.3). The sample preparation procedure is depicted in Fig. 6.2, where the Dynabead-conjugated cells were suspended in a methyl cellulose solution (a viscous fluid) and were subjected to alternating magnetic fields with a period of 1 s and a square-wave driving current. As shown in Figs. 6.3a-o, due to the high viscosity of the methyl cellulose solution, the labelled cells mainly demonstrated 3D rotational motion. Typically, the motion of a labelled cell starts at the beginning of a cycle of the magnetic field (e.g., $t = 0.5$ s), approaching a steady state (e.g., $t = 1.0$ s) before the magnetic field switches its direction and the cell rotates in the reverse direction (e.g., between $t = 1.0$ s and $t = 1.5$ s). The two extreme positions of the rotational motion are demonstrated in Fig. 6.3p by overlaying the images captured at $t = 0.5$ s and $t = 1.0$ s using magenta and green, respectively.

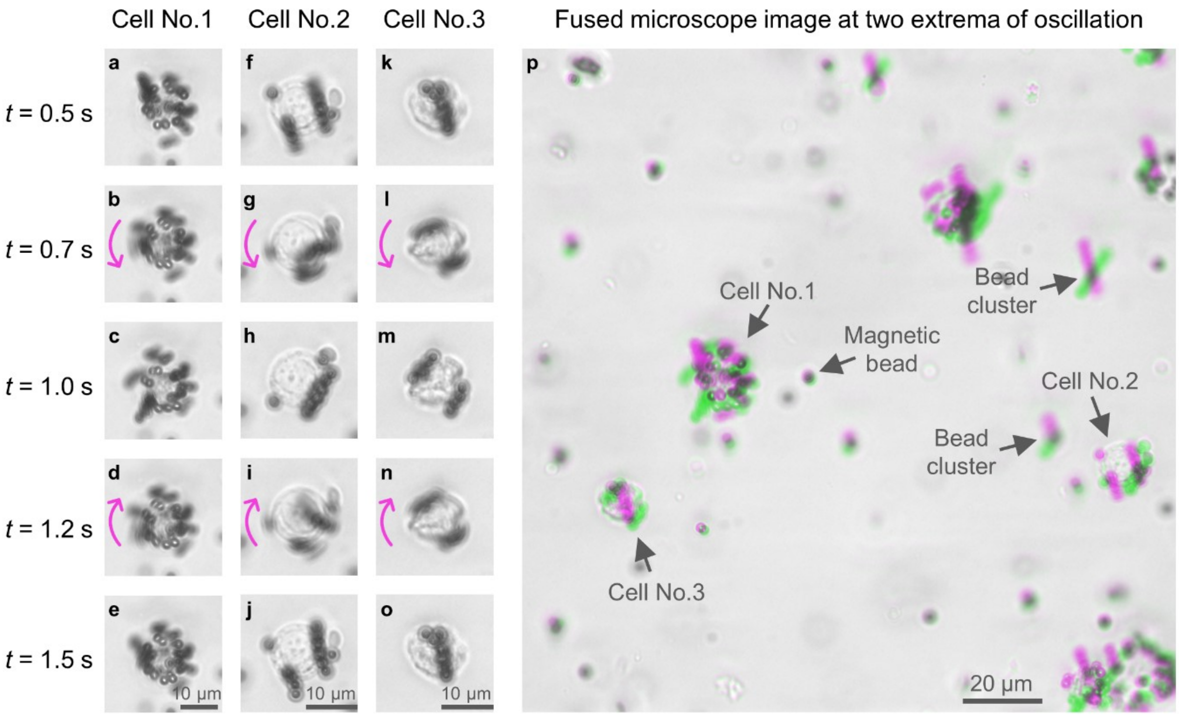


Figure 6.3 Dynabead-conjugated MCF7 cells demonstrate periodic rotational motion under an alternating magnetic force field. Images were acquired using a 40× 0.6NA benchtop microscope. a-o Snapshots of three Dynabead-conjugated MCF7 cells at different time points within a period of oscillation (period = 1 s). p Images taken at the two extrema of the oscillation (*t* = 0.5 s and *t* = 1.0 s) were fused together to demonstrate the movement, where the grey regions in the fused image represent the consistency between the two images and the magenta/green colours represent the differences of the two images. Magenta represents the first image (*t* = 0.5 s), and green represents the second image (*t* = 1.0 s).

Various unbound magnetic beads and bead clusters are also observed within the sample (Fig. 6.3p reports some examples, marked with text and arrows), which also oscillate at the same frequency as that of the bead-conjugated target cells. If not handled properly, these might form a major cause of false positives. However, the spatio-temporal dynamics of bead-conjugated cells significantly differ from those of unbound beads and bead clusters (see the following subsections and the Methods section). For a given amount of magnetic driving force, the bead-conjugated cells are subjected to more inertia and viscous drag, which is manifested by a slower response to the magnetic field, i.e., a slower rotational motion. In addition, magnetic beads typically form chains when they cluster under an external magnetic field, and these chains exhibit a swinging motion under the alternating magnetic field. This contrasts with the 3D rotational motion, i.e., the “rolling” motion associated with the bead-conjugated cells. These intricate spatio-temporal dynamic features, in addition to morphological differences, are utilized by a subsequent classification step (based on a deep neural network) to achieve higher accuracy and eliminate false positive detections, as will be detailed in the following subsections and the Methods section.

Cell detection and classification using CMA and deep learning

The sample, which contains the periodically oscillating target cells and other types of unwanted background particles, is illuminated with coherent light. The interference pattern recorded by the CMOS image sensor represents an in-line hologram of the target cells, which is partially obscured by the random speckle noise resulting from the background particles, including other unlabelled cells, cell debris and unbound magnetic particles. Recorded at 26.7 frames per second using the CMOS image sensor, these patterns exhibit spatio-temporal variations that are partially due to the controlled cell motion. This phenomenon is exploited for the rapid detection of magnetic-bead-conjugated rare cells from a highly complex and noisy background. Figs. 6.4a-g show the detailed computational steps for the preliminary screening of cell candidates from a raw holographic image sequence. First, a computational drift correction step mitigates the overall drift of the sample between frames. Then, a high-pass filtered back-propagation step using the angular spectrum method [201] calculates the holographic images at different axial distances within the 3D sample. A CMA step analyses the differences among the frames to enhance the 3D contrast for periodically moving objects that oscillate at the driving frequency and employs time averaging to suppress the random speckle noise caused by background particles. This is then followed by a maximum intensity projection and threshold-based detection to locate potential cell candidates.

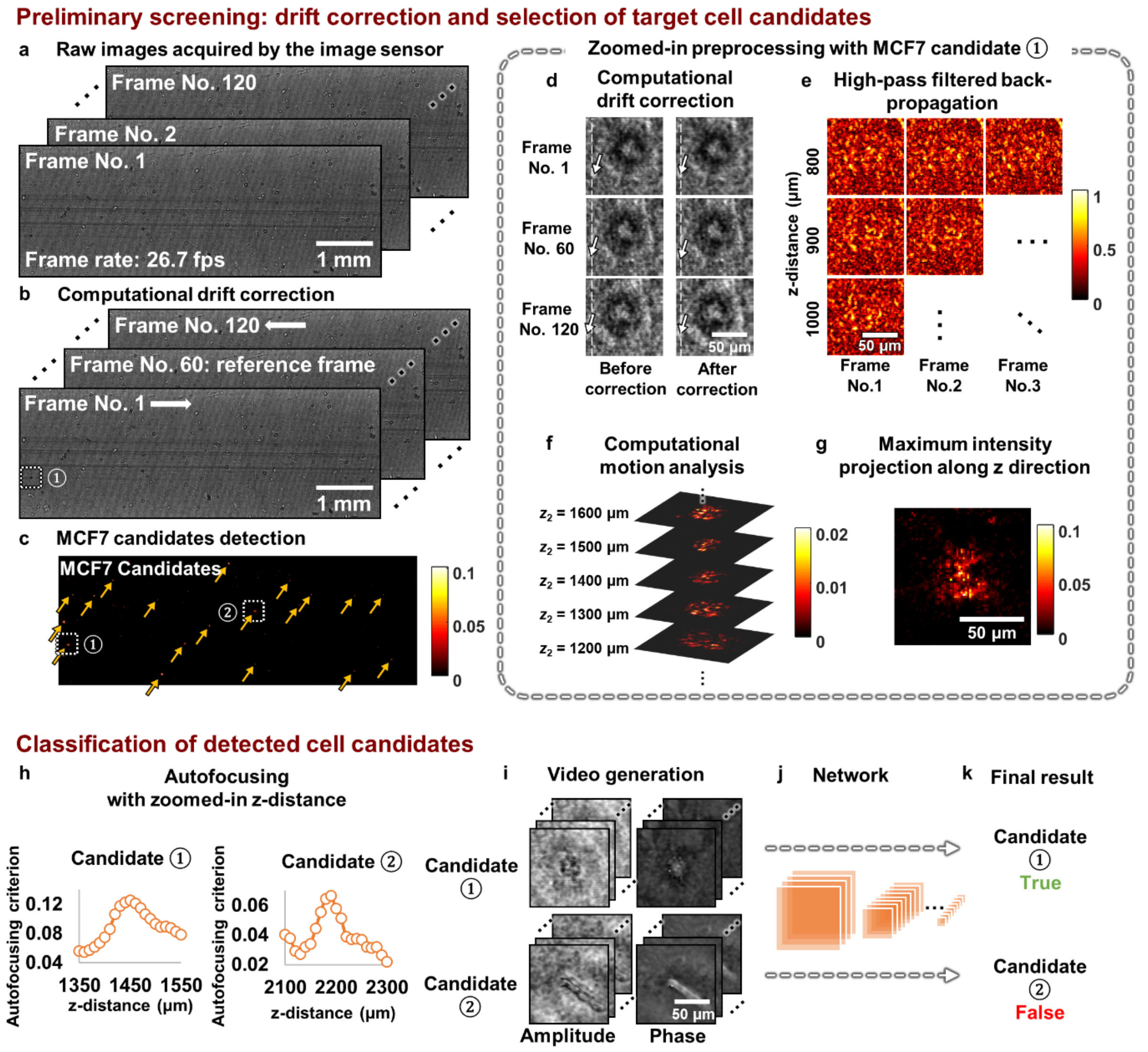


Figure 6.4 Computational detection of rare cells. a-c Preliminary screening of the whole FOV to detect candidates for target cells (MCF7). At each scanning position, 120 frames of raw holograms were taken at 26.7 frames per second. Computational drift correction was applied to mitigate the horizontal shift caused by the fluid drift, where the vertical movement caused by the magnetic field was kept unmodified. The lateral position of each MCF7 candidate was located by CMA, maximum intensity projection and threshold-based detection. **d-g** Zoomed-in preliminary processing for the example region labelled ① in **b-c**. **h-k** Classification process for the two cell candidates labelled ① and ② in **c**. The axial location for each cell candidate was determined by autofocusing. A video was formed for each cell candidate by propagating each frame to the in-focus plane. The classification was performed by a densely connected P3D convolutional neural network, as detailed in the Methods section.

The cell candidates that are detected in this preliminary screening step contain a large number of false positives, which mainly result from unbound magnetic beads that form clusters under the external magnetic field. Therefore, we employ another classification step (Fig. 6.4h-k) to improve the specificity of our final detection. For this classification step, we choose to use a densely connected P3D CNN structure to classify the holographic videos to exploit the spatial and temporal information encoded in the captured image sequence. The densely connected P3D CNN structure is modified based on a recently proposed CNN structure [200] by adding dense connections. [202] Compared to other machine learning techniques, the use of a deep neural network for video classification is typically more powerful, and the network can be retrained to classify other types of cells or objects of interest. [200,203]

An autofocusing step [70,204] is applied to each candidate object to create an in-focus amplitude and phase video, which is then classified (as positive/negative) by a densely connected P3D CNN. These classification results are used to generate the final rare cell detection decisions and cell concentration measurements. The CNN was trained and validated with manually labelled video clips generated from ten samples that were used solely for creating the training/validation datasets. This training needs to be performed only once for a given type of cell-bead conjugate (for details, refer to the Methods section).

Evaluation of system performance

To quantify the LoD of our platform for detecting MCF7 cells in human blood, we spiked cultured MCF7 cells in whole blood at various concentrations and used our technique to detect the spiked MCF7 cells. Using spiked samples instead of clinical samples provides a well-defined system to characterize and quantify the capabilities of our platform, which is an important step before moving to clinical samples in the future.

In each experiment, 4 mL of MCF7-spiked whole human blood at the desired concentration was prepared. Then, the procedure in Fig. 6.2 was followed to perform magnetic separation and embed the recovered cells in the viscous methyl cellulose medium, resulting in ~3.2 mL of final sample volume. This prepared sample was then loaded into a disposable capillary tube to be screened by our computational cytometer. Because the capillary tube length is designed to be longer than the range of the motion of the linear stage and because the capillary tube was wider than the width of the CMOS sensor, the actual imaged volume per test (within the sample tube) is ~0.942 mL, which corresponds to ~1.177 mL of the blood sample before the enrichment process.

MCF7 concentrations of 0 mL^{-1} (negative control), 10 mL^{-1}, 100 mL^{-1} and 1000 mL^{-1} were tested, where three samples for each concentration were prepared and independently measured. Fig. 6.5 shows the results of the blind testing of our technique using serial dilution experiments. The blue data points correspond to a one-time testing result, where the error bars correspond to the standard deviations of the three detected concentrations at each spiked concentration. Without the detection of any false positives in the negative control samples, our technique was able to consistently detect MCF7 cells from 10 mL^{-1} samples, measuring a target cell concentration of 1.98 ± 1.06 mL^{-1}. At this low concentration (10 cells/mL), the detection rate was approximately 20%. The experimentally measured detection rate dropped to ~5% at a higher concentration of 1000 cells/mL.

Because the training of the deep neural network inherently includes randomness, we further evaluated the repeatability of our network training process. For this, we randomly and equally divided our training data into five subsets, and then we trained five individual networks by assigning one different subset as the validation dataset and the combination of the remaining four subsets as the training dataset. Each of the five networks was blind tested to generate the serial dilution results. The mean and standard deviation of the detected concentrations resulting from the five networks are shown in Fig. 6.5 (orange data points; for each trained network, three detected concentrations are averaged at each spiked concentration). Overall, good consistency between the different network results is observed.

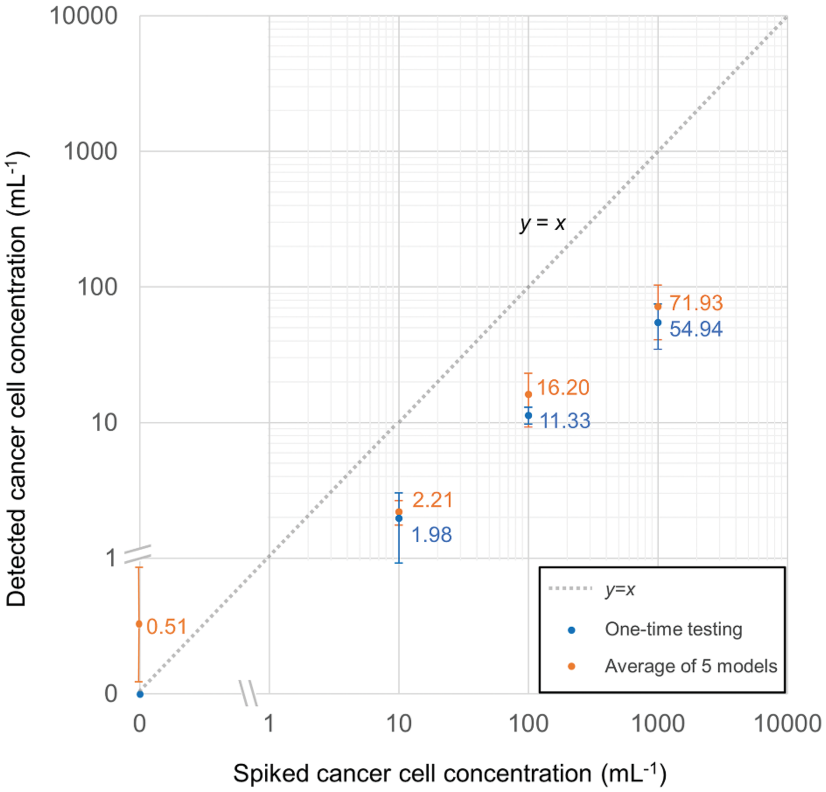


Figure 6.5 Quantification of the LoD of our computational cytometer based on magnetically modulated lensless speckle imaging for the detection of MCF7 cells in whole blood. The axes are a hybrid of logarithmic and linear scales to permit 0 cells/mL to be shown in the same plot. The blue data points represent one-time testing results of a single trained P3D CNN. The error bars represent the respective standard deviation of the three repeated tests at each spiked target cell concentration. The orange data points represent the averaged testing results using five P3D CNNs that were individually trained on a different subset of data. The error bars represent the standard deviation resulting from the detections of the five individual networks; for each trained network, three detected concentrations are averaged at each spiked concentration.

The underdetection behaviour of our system is due to a combination of both systematic errors and random factors. A major reason for underdetection is the tuning of the classification network. In the *preliminary* screening step, because there are typically a large number of false positive detections and a low number of true positive detections (since the target cells are quite rare), our classifier must be tuned to have an extremely low false positive rate (FPR) to have a low LoD. To satisfy this, we applied a widely adopted method for tuning our classifier, [205] where we selected a decision threshold based on the training/validation dataset, which leads to a zero FPR (see the Methods section for details). However, an inevitable side effect of reducing the FPR is a reduction in the true positive rate (TPR). Based on the validation results, when a decision threshold of 0.999999 was used, the TPR dropped to 10.5%. This explains a major portion of the reduced detection rate that we observed in the serial dilution tests (Fig. 6.5). Another systematic error that contributes to the underdetection is the imperfect recovery rate of MCF7 cells during the enrichment. We experimentally quantified the recovery rate of MCF7 cells using Dynabeads to be ~85% (Table S6.1).

The remainder of the underdetection and fluctuations in the detection rate at different concentrations may be associated with various other factors, e.g., sample handling errors (especially at low cell concentrations), clustering of the target cells, and non-uniform labelling of cells with magnetic beads. In fact, MCF7 cells are known to form clusters and have thus been extensively used for preparing *in vitro* tumour models. [206,207] In an experiment where we spiked MCF7 cells at a concentration of 1.1×10^5/mL (Table S6.1), we observed that ~50% of the MCF7 cells formed clusters after enrichment. However, the amount of clustering is expected to be lower at decreased MCF7 concentrations, which partially explains our reduced detection efficiency at higher cell concentrations. This clustering of cells not only reduces the overall number of target entities but may also exhibit changes in their oscillation patterns and may be misclassified by our classifier.

6.3 Discussion

The presented computational cytometry technique may be applied for the detection of various types of rare cells in blood or other bodily fluids using appropriately selected ligand-coated magnetic beads. There are several advantages of our magnetically modulated speckle imaging technique. The first important advantage is its ability to detect target rare cells without any additional modification, such as labelling with fluorescent or radioactive compounds, unlike the vast majority of the existing techniques. The same magnetic beads that are used for capturing and isolating target cells from whole blood are also used for periodic cell modulation and specific detection within a dense background. False positives are mitigated by identifying the controlled spatio-temporal patterns associated with the labelled target cells through a trained deep neural network.

Compared to existing approaches, our technique also has the advantages of a relatively low LoD, rapid detection and low cost, which makes it suitable for the sensitive detection of rare cells in resource-limited settings. For example, fluorescence imaging and Raman microscopy have been widely used to detect rare cells and have been shown to have very low LoDs (e.g., ~1 cell/mL), [159,208,209] but they are typically limited by a high system cost and complexity. To address this issue, a low-cost fluorescence system for detecting rare cells was introduced by Balsam *et al.*, [210] which detects fluorescently labelled cells flowing in a fluidic channel using laser excitation and a low-cost camera for imaging. They demonstrated an LoD comparable to ours (~10 mL^{-1}) for SYTO-9-labelled THP-1 monocytes in whole blood. However, the use of fluorescence labelling can suffer from the drawback of photobleaching. As another notable example for sensitive and cost-effective rare cell detection, Issadore *et al.* proposed using Hall sensors to detect magnetic-bead-labelled target cells in a microfluidic channel and demonstrated a high sensitivity in detecting CTCs. [211] However, their technique requires a relatively long detection time (2.5 hours) and a strong expression of biomarkers in target cells. Other rare cell detection technologies, such as chemiluminescence detection based on aptamer-specific cell capture [212] and DNA-oriented shaping of cell features [213], have also been reported, but their capabilities were demonstrated using only cell mixtures in a buffer solution with limited throughput, i.e., 3 μL [212] or 500 μL [213] cell solution per batch.

In our approach, while deep-learning-based classification is instrumental to achieving high detection accuracy, it needs to be retrained on different types of cells, which requires collecting and labelling a large amount of data for each new type of target cell. This is a disadvantage of our approach; however, preparing the training data and manually labelling the target cells is not prohibitively time consuming, and it needs to be performed only once –, i.e., during the training phase. For example, when we prepared the training/validation data for MCF7 cells, we used 10 experiments to create a manually labelled library containing 17,447 videos of candidate cells (including positives and negatives). The manual labelling process took approximately 10 hours. These procedures only need to be performed once for a given type of target cell. Compared to using fluorescence labelling, which requires additional experimental steps and reagents each time, we believe that this one-time cost of preparing training data for the deep neural network presents advantages.

Another limitation of our method is that it can detect only positive cells, which are labelled with magnetic beads; negative cells that are not labelled are not counted. Additionally, in this proof-of-concept study, we only demonstrated our detection technique on a single type of target cell. However, a future direction would be to explore the feasibility of multiplexed labelling for different types of target cells. One possibility for multiplexing is to use magnetic particles of different sizes (e.g., varying from ~100 nm – 10 μm), shapes and iron content, where each type of magnetic particle is coated with the corresponding antibody that is specific to a different type of cell. In this approach, different cell-bead conjugates would have distinct dynamics when they are subjected to a varying magnetic force field, which would lead to different patterns of oscillation that can be specifically detected. [194,195] The cell-bead conjugates may also exhibit different responses to magnetic modulation when the frequency is varied. These spatio-temporal and morphological signatures may be classified by an appropriately designed and trained deep-learning-based classifier. Therefore, any type of rare cell that can be specifically identified/isolated using antibodies or any targeting moieties can potentially be targeted using our presented system.

The spatial resolution and the quality of the images captured using our system are degraded by the random speckle noise generated by background objects, which limits our ability to perform further morphological analysis based on reconstructed images. However, at different frames of a video that is captured with our system, since the target objects of interest (i.e., the bead-labelled MCF7 cells) are modulated with a unique spatio-temporal pattern, exposing different perspectives of the cells, a robust distinction of the target cells from the background is achieved using our deep-learning-based video classifier.

The entire prototype of our computational cytometer shown in Fig. 6.1b (excluding the function generator, power supply and laptop computer) has a raw material cost of ~$750. This cost can be significantly reduced under large volume manufacturing, and it is currently mainly attributed to the image sensor and frame grabber (~$550), the permalloy rod (~$70), and the electromagnets (~$40), with the other components being much more inexpensive. In future versions of this instrument, the power supply and function generator can be replaced with cost-effective integrated circuit chips. For example, the power supply can be replaced with a 20 V power adapter (e.g., TR9KZ900T00-IMR6B, GlobTek, Inc., Northvale, NJ, USA) and a step-down converter (e.g., LTC3630EMSE#PBF, Analog Devices, Norwood, MA, USA) to generate 20 V and 12 V power for the electromagnets and the stepper motor, respectively; the function generator can be replaced with an oscillator circuit built from a timer integrated circuit (e.g., NE555DR, Texas Instruments, Dallas, TX, USA). The total cost of these components would be less than $25. Furthermore, the device can be easily scaled up to include two or more parallel imaging channels to achieve a higher sample throughput, which is proportionate with the number of imaging channels.

6.4 Materials and methods

Cell preparation

MCF7 cell lines were purchased from ATCC (Manassas, Virginia, USA). Cells were plated with 10 mL of growth media in a T75 flask (Corning Inc., New York, USA) at a concentration of 1×10^5 cells/mL. The growth media was composed of Dulbecco's Modified Eagle Medium (DMEM, Gibco®, Life Technologies, Carlsbad, California, USA) supplemented with 10% (v/v) foetal bovine serum (FBS, Gibco®, Life Technologies, Carlsbad, California, USA) and 1% penicillin-streptomycin (Sigma-Aldrich Co., St. Louis, Missouri, USA). Cells were grown in a humidified incubator at 37°C in a 5% CO_2 environment. Cells were harvested by treating them with 0.25% trypsin-edta (Gibco®, Life Technologies, Carlsbad, California, USA) for 3 min 2-3 days after seeding, depending on confluency. Then, the cells were pelleted by centrifuging for 3 min at 1200 RPM and resuspended in the growth media to a final concentration of 1×10^6 cells/mL.

Sample preparation

Rare cell dilution: The MCF7 cells were serially diluted in Dulbecco's phosphate-buffered saline (DPBS, Sigma-Aldrich Co., St. Louis, Missouri, USA) at different concentrations (2×10^4 cells/mL, 2×10^3 cells/mL, and 2×10^2 cells/mL). The dilution of MCF7 cells in whole blood was prepared by mixing the cell solution with whole blood at a ratio of 1:19 (v/v). Most of the experiments were performed by mixing 200 μL of cell solution with 3.8 mL of whole blood. Healthy human whole blood (from anonymous and existing samples) was obtained from the UCLA Blood and Platelet Center.

Bead washing: CELLection Epithelial Enrich Dynabeads (Invitrogen, Carlsbad, California, USA) were first resuspended in DPBS and vortexed for 30 s. A magnet (DX08B-N52, K&J

Magnetics, Inc., Pipersville, Pennsylvania, USA) was then used to separate the Dynabeads, and the supernatant was discarded. This process was repeated three times, and the Dynabeads were resuspended in DPBS at the initial volume.

Rare cell separation: The washed Dynabeads were then added to the MCF7-spiked whole blood sample at a concentration of 2.5 µL beads per 1.0 mL of blood sample. The mixture was incubated for 30 min with gentle tilting and rotation. A magnet was placed under the vial for 5 min, and the supernatant was discarded after that. To this solution, we added 1 mL of cold DPBS buffer and mixed it gently by tilting from side to side. This magnetic separation procedure was repeated five times. After the final step, the sample was resuspended in 0.7 mL of DPBS and gently mixed with 2.5 mL of 400 cP methyl cellulose solution (Sigma-Aldrich Co., St. Louis, Missouri, USA) using a pipette. The sample was incubated for 5 min to reduce the number of bubbles before it was loaded into a glass capillary tube (Part # BRT 2-4-50; cross-section inner dimension of 2 mm×4 mm; $11.80 per foot; Friedrich & Dimmock, Inc., Millville, New Jersey, USA). The ends of the capillary tube were sealed with parafilm before the tube was mounted onto our computational cytometer for imaging and cell screening.

Design of the computational cytometer based on magnetically modulated lensless speckle imaging

As shown in Fig. 6.1, our device hardware consists of an imaging module and a linear translation module. The imaging module, i.e., the scanning head in Fig. 6.1, contains a laser diode (650 nm wavelength, AML-N056-650001-01, Arima Lasers Corp., Taoyuan, Taiwan) for illumination, which has an output power of ~1 mW. The sample is loaded inside a capillary tube with a rectangular cross section, which is placed ~7.6 cm below the light source. A CMOS image sensor (acA3800-14um, Basler, Ahrensburg, Germany) with a pixel size of 1.67 µm, which is placed below the glass tube with a narrow gap (~1 mm), is used to capture the holographic speckle patterns generated by the liquid sample. To induce oscillatory motion to the labelled cells in the sample, two electromagnets (Part #XRN-XP30×22, Xuan Rui Ning Co., Ltd., Leqing, Zhejiang Province, China) with custom-machined permalloy extensions are placed on either side of the glass tube. An alternating driving current (square wave) is supplied to either of the electromagnets, with a 180° phase shift between them, which creates alternative pulling force to the magnetic particles within the sample. The low level of the driving current is 0, and the high level of the driving current is ~500 mA. The frequency is 1 Hz, which was experimentally optimized to maximize the signal corresponding to the magnetic-bead-conjugated cancer cells.

The linear translation stage is custom-built using off-the-shelf components. A bipolar stepper motor (No. 324, Adafruit Industries LLC., New York, USA) with two timing pulleys and a timing belt is used to provide mechanical actuation, and the imaging module is guided by a pair of linear motion sliders and linear motion shafts on either side of the scanning head. 3D-printed plastic is used to construct the housing for the scanning head, and laser-cut acrylic is used to create the outer shell of the device.

Image acquisition

After the sample is loaded into the capillary tube and placed onto our computational cytometer, the image acquisition procedure begins. The linear translation stage moves the scanning head to a series of discrete positions along the glass tube. At each position, the stage stops, allowing the CMOS image sensor to capture a sequence of 120 holograms at a frame rate of 26.7 fps before moving onto the next position. The image data are saved to a solid-state drive (SSD) for storage and further processing.

Because the FOV corresponding to the edges (i.e., top and bottom rows) of the image sensor is subject to a highly unbalanced magnetic force field due to the closeness to one of the electromagnets, only the central 1374 rows of the image sensor's pixels are used to capture the image sequence, where the magnetic force from the two electromagnets are relatively balanced.

Because the CMOS image sensor temperature quickly rises when it is turned on, it tends to cause undesired flow inside the glass tube due to convection. Therefore, a scanning pattern is engineered to reduce the local heating of the sample: if we denote 1, 2, …, 32 as the indices of the spatially adjacent scanning positions, the scanning pattern follows 1, 9, 17, 25, 2, 10, 18, 26, …. This scanning pattern ensures that a given part of the sample cools down before the scanning head moves back to its neighbourhood. The power to the image sensor is also cut off during the transition between the two successive scanning positions, which is implemented by inserting a MOSFET-based switch into the power line of the USB cable.

Computational detection and localization of cell candidates and deep-learning-based classification

The image processing procedure (Fig. 6.4) can be divided into two parts: (1) a preliminary screening step, which applies computational drift correction and MCF7 candidate detection to the entire FOV to locate target cell candidates in 2D, and (2) a classification step, which refocuses the holographic image sequence to each individual MCF7 candidate in its local area, generates an in-focus amplitude and phase video for each candidate, and classifies the corresponding video with a trained deep neural network. This procedure is further detailed below.

1. Preliminary screening

Computational drift correction

The sample fluid in the glass capillary tube often drifts slowly throughout the duration of the image acquisition, which is due to, e.g., the imperfect sealing at the ends of the tube and the convection due to the heat of the image sensor. Because the detection and classification of the target cells are largely based on their periodic motion, the drifting problem should be corrected. Since our sample is embedded within a viscous methyl cellulose, minimal turbulent flow is observed, and the drifting motion within our imaged FOV is almost purely translational. We used a phase correlation method [214] to estimate the relative translation between each frame in the sequence with respect to a reference frame (chosen to be the middle frame in the holographic image sequence) and used 2D bilinear interpolation to remove the drift between frames. As shown in Fig. S6.2, this drift correction step successfully removed many false positive detections in the CMA step due to the background drift.

Detection of target cell candidates

The detection of the target cell *candidates* plays a key role in automatically analysing the sample, as it greatly narrows down the search space for the rare cells of interest and allows the subsequent deep-learning-based classification to be applied to a limited number of holographic videos. In the preliminary screening stage, the lateral locations of the MCF7 candidates are detected. Each frame of the raw hologram sequence is propagated to various axial distances throughout the sample volume using a high-pass-filtered angular spectrum propagation kernel, which can be written as:

$$\mathbf{B}_i\left(z_j\right) = HP\left[\mathcal{P}\left(\mathbf{A}_i, z_j\right)\right] \tag{56}$$

where $HP(\cdot)$ denotes the high-pass filter (see Supplementary Information for details), $\mathcal{P}(\cdot)$ denotes angular spectrum propagation, [201] $\mathbf{A}_i$ denotes the i-th frame of the raw hologram sequence after the drift correction, and z_j denotes the j-th propagation (axial) distance. The selected propagation distances ranged from 800 μm to 5000 μm with a step size of 100 μm to ensure coverage of all possible MCF7 candidates within the sample tube. A zoomed-in image of $\mathbf{B}_i(z_j)$ corresponding to an example region is shown in Fig. 6.4e.

Next, for every given propagation distance, a CMA algorithm is applied to reveal the oscillatory motion of the target cells within the sample, which focuses on periodic changes in the recorded frames:

$$\mathbf{C}\left(z_j\right)=\frac{1}{N_{\mathrm{F}}-N}\sum_{i=1}^{N_{\mathrm{F}}-N}\left(\frac{1}{2}\left|\mathbf{B}_i\left(z_j\right)-\mathbf{B}_{i+N/2}\left(z_j\right)\right|+\frac{1}{2}\left|\mathbf{B}_{i+N/2}\left(z_j\right)-\mathbf{B}_{i+N}\left(z_j\right)\right|-\left|\mathbf{B}_i\left(z_j\right)-\mathbf{B}_{i+N}\left(z_j\right)\right|\right) \quad (57)$$

where $\mathbf{C}(z)$ and $\mathbf{B}(z)$ are shorthand notations for $\mathbf{C}(x, y; z)$ and $\mathbf{B}(x, y; z)$, respectively, N_{F} is the total number of recorded frames (in our case, $N_{\mathrm{F}} = 120$), and N is chosen such that the time difference between the i^{th} frame and the $(i+N)^{\text{th}}$ frame is equal to the period of the alternating magnetic field. Therefore, the first two terms inside the summation in Equation (54) represent half-period movements at the j-th propagation distance, and the last term represents the whole-period movement. Ideally, for objects that oscillate periodically with the alternating magnetic force field, the first two terms should be relatively large, and the last term should be relatively small. For randomly moving objects, the three terms in the summation approximately cancel each other out. As a result, $\mathbf{C}(x, y; z)$ is a 3D contrast map that has high values corresponding to the locations of periodic motion that match the frequency of the external magnetic field. An example of **C** is shown in Fig. 6.4f.

To simplify segmentation, a maximum intensity projection along the axial direction (i.e., z) is applied to flatten the 3D image stack into a 2D image, which can be written as:

$$\mathbf{D}\left(x,y\right)=\max_{z}\left[\mathbf{C}\left(x,y;z_1\right),\mathbf{C}\left(x,y;z_2\right),\ldots,\mathbf{C}\left(x,y;z_{N_{\mathrm{H}}}\right)\right] \quad (58)$$

where x and y are the lateral indices and N_{H} is the total number of axial positions (in our case, $N_{\mathrm{H}} = 43$). An example of **D** is shown in Fig. 6.4c, with a zoomed-in image shown in Fig. 6.4g. Thresholding-based segmentation was applied to the calculated 2D image **D**, and the resulting centroids are used as the lateral positions of the MCF7 candidates.

2. Classification

Autofocusing and video generation

After the preliminary screening, which identifies the lateral centroids of potential target cell candidates, the subsequent processing is applied to each MCF7 candidate only within their local area. Autofocusing [70,204] was first performed to locate the MCF7 candidate in the axial direction. Because $\mathbf{C}(x, y; z_j)$ should have a higher value when approaching the in-focus position of each MCF7 candidate, the approximate axial position was obtained by maximizing (as a function of z_j) the sum of the pixel values of $\mathbf{C}(x, y; z_j)$ ($j = 1, 2, \ldots, N_\mathrm{H}$) in a local neighbourhood around each individual MCF7 candidate. We chose to use a local neighbourhood size of 40×40 pixels (i.e., 66.8 µm×66.8 µm). This process can be written as follows:

$$\hat{z}_k = \underset{z_j=1,2,\ldots,N_H}{\arg\max} \sum\nolimits_{x,y=-19}^{20} \mathbf{C}\left(x_k + x, y_k + y; z_j\right) \tag{59}$$

where $\hat{z}_k$ is the resulting in-focus position for the k-th potential target cell candidate and x_k and y_k are the lateral centroid coordinates of the k-th potential target cell candidate.

The same criterion to find the focus plane can be applied again with finer axial resolution to obtain a more accurate estimation of the axial distance for each MCF7 candidate. We used a step size of 10 µm in this refined autofocusing step. Two examples of this process are shown in Fig. 6.4h. Alternatively, the Tamura coefficient [70,204] could also be used as the autofocusing criterion to determine the in-focus plane.

Finally, the in-focus amplitude and phase video corresponding to each MCF7 candidate was generated by digitally propagating every frame of the drift-corrected hologram sequence to the candidate's in-focus plane. The final video has 120 frames at 26.67 fps with both the amplitude and phase channels, and each frame has a size of 64×64 pixels (pixel size = 1.67 µm). Two examples corresponding to two cell candidates are shown in Fig. 6.4i.

Target cell detection using densely connected P3D CNN

Each video of the MCF7 candidate was fed into a classification neural network (Fig. 6.6), which outputs the probability of having an MCF7 cell in the corresponding video (Fig. 6.4j-k). We designed a novel structure for the classification neural network, named densely connected P3D CNN, which is inspired by the pseudo-3D residual network [200] and the densely connected convolutional network. [202] The original P3D CNN [200] used a mixture of three different designs of the P3D blocks to gain structural diversity, which resulted in a better performance. In this work, we introduced a densely connected structure to the P3D CNN structure by adding dense (skip) connections inside the spatio-temporal convolution block (dashed black arrows in Fig. 6.6 inset) to unify the three different P3D blocks. This allowed a simpler network design that was easier to implement for our task.

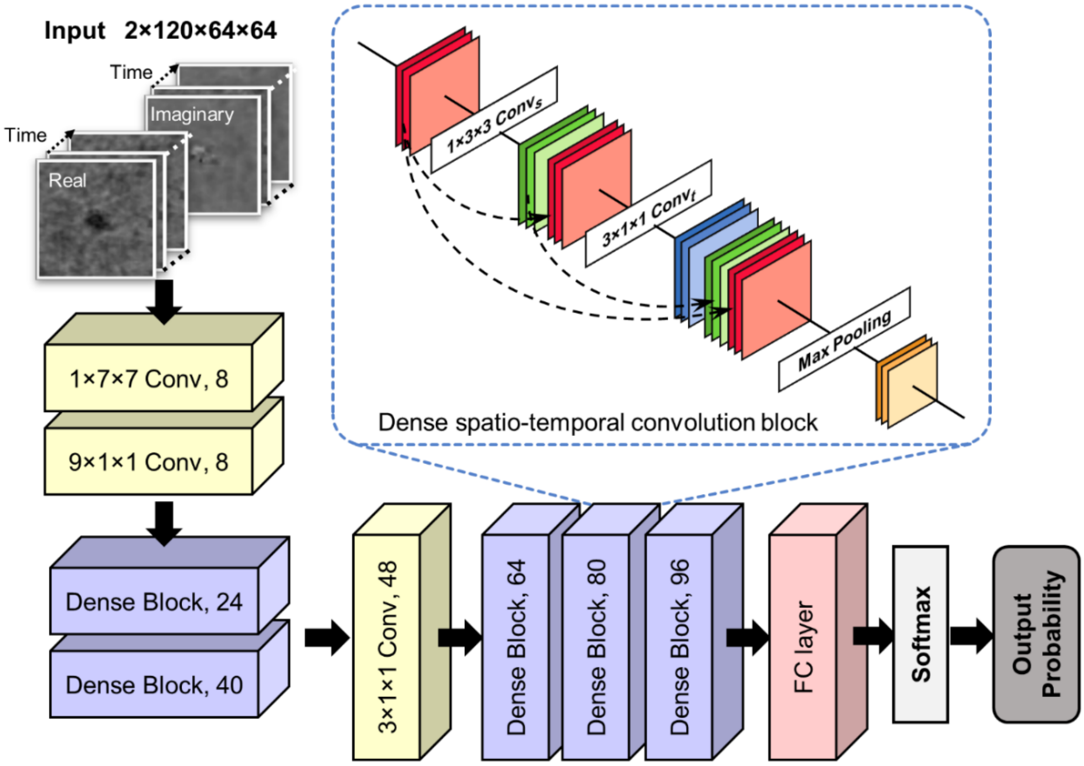


Figure 6.6 Structure of the densely connected P3D CNN. The network consists of convolutional layers, a series of dense blocks, a fully connected layer and a softmax layer. As shown in the inset, each dense spatio-temporal convolution block was constructed by introducing skip connections between the input and output of

the convolutional layers in the channel dimension, where red represents the input of the dense block, green and blue represent the output of the spatial and temporal convolutional layers, respectively, and yellow represents the output of the entire block.

The detailed structure of the densely connected P3D CNN is shown in Fig. 6.6. The network contains five densely connected spatio-temporal convolutional blocks. As shown in the inset of Fig. 6.6, each block consists of a 1×3×3 spatial convolutional layer (Conv_s), a 3×1×1 temporal convolutional layer (Conv_t), and a max pooling layer (Max). Each spatial (or temporal) convolutional layer is a composition of three consecutive operations: batch normalization, a rectified linear unit (ReLU) and a spatial (or temporal) convolution (with stride = 1 and output channel number equal to the growth rate $k = 8$). In each block, we introduced skip connections between the input and output of the Conv_s layer as well as the Conv_t layer by concatenating ($\oplus$) the input and the output in the channel dimensions. For a given input tensor m_p, the densely connected spatio-temporal convolutional block maps it to the output tensor m_{p+1}, which is given by:

$$m_{p+1} = \mathrm{Max}\left[\mathrm{Conv}_t\left(\mathrm{Conv}_s(m_p) \oplus m_p\right) \oplus \left(\mathrm{Conv}_s(m_p) \oplus m_p\right)\right] \tag{60}$$

For example, consider an input video with a size of $c \times t \times h \times w$, where c, t, h and w denote the number of channels, number of frames (time), height and width of each frame (space), respectively. Here, $c = 2$, $t = 120$, and $h = w = 64$. We first pass the video through a 1×7×7 spatial convolutional layer (stride = 2) and a 9×1×1 temporal convolution layer (stride = 3) sequentially. The output channel numbers of the layers are included in Fig. 6.6 in each box. Then, the data go through 5 dense blocks, where between the 2nd and 3rd dense blocks, we add an additional 3×1×1 (stride = 1) convolutional filter with no padding to ensure that the time and space dimensions are equal. A fully connected (FC) layer with a 0.5 dropout rate and a softmax layer are introduced, which output the class probability (target rare cell or not) for the corresponding input video. Finally, a decision threshold is applied to the class probability output to generate the final positive/negative classification, where the decision threshold is tuned based on the training/validation data to reduce the FPR (detailed in the next sub-section, *Network training and validation*).

Network training and validation

We performed ten experiments (i.e., ten samples) to create the training/validation datasets for our classifier and then used the trained classifier to perform blind testing on additional serial dilution experimental data (Fig. 6.5), which had no overlap with the training/validation data. Among the ten experiments for constructing the training/validation dataset, 5 were negative controls, and the other 5 were spiked whole blood samples at an MCF7 concentration of 10^3 mL^{-1}. When manually labelling the video clips to create the training/validation dataset, we noticed that some videos were difficult to label, where the annotators could not make a confident distinction. Therefore, to ensure an optimal labelling accuracy, our negative training data came from only the 5 negative control experiments, where all the candidate videos from those experiments were used to construct the negative dataset. The positive training data were manually labelled by two human annotators using 5 experiments spiked at 10^3 mL^{-1}, where only the video clips that were labelled as positive with high confidence by both annotators were selected to enter the positive training dataset, while all the others were discarded.

Next, the training/validation datasets were randomly partitioned into a training set and a validation set with no overlap between the two. The training set contained 1713 positive videos and 11324 negative videos. The validation set contained 788 positive videos and 3622 negative videos. The training dataset was further augmented by randomly mirroring and rotating the frames by 90°, 180° and 270°. The convolutional layer weights were initialized using a truncated normal distribution, while the weights for the FC layer were initialized to zero. Trainable parameters were optimized using an adaptive moment estimation (Adam) optimizer with a learning rate of 10^{-4} and a batch size of 240. The network converged after ~800-1000 epochs. The network structure and hyperparameters were first optimized to achieve high sensitivity and specificity for the validation set. At a default decision threshold of 0.5, a sensitivity and specificity of 78.4% and 99.4%, respectively, were achieved for the validation set; a sensitivity and specificity of 77.3% and 99.5%, respectively, were achieved for the training set. After this initial step, because our rare cell detection application requires the classifier to have a very low FPR, we further tuned the decision threshold of our classifier to avoid false positives. For this, the training and validation datasets were combined to increase the total number of examples, and we gradually increased the decision threshold (for positive classification) from 0.5 while monitoring the FPR for the combined training/validation dataset. We found that a decision threshold of 0.99999 was able to eliminate all false positive detections in the combined training/validation dataset. We further raised the decision threshold to 0.999999 to account for potential overfitting of the network to the training/validation data and further reduced the risk of false positive detections.

At a decision threshold of 0.999999, as expected, the TPR dropped down to 10.5% (refer to Fig. S6.3, which reports the receiver operating characteristic (ROC) curve based on the validation dataset, with an area under the curve of 0.9678). This low TPR results in underdetection of the target cells, as also evident in our serial dilution results (Fig. 6.5). The selection of the decision threshold is dependent on the specific application of interest and should be tuned based on the expected abundance of target cells and the desired LoD. For the application considered in this work, because the expected number of target cells at the lowest concentration (i.e., 10 mL^{-1}) is extremely low, the decision threshold was tuned to a high level to suppress false positives, which in turn resulted in a very low TPR. However, for less demanding cell detection or cytometry applications where the desired LoD is not as stringent, the decision threshold may be relaxed to a lower level, which also allows the TPR to be higher.

Computation time

Using our current computer code, which is not optimized, it takes ~80 s to preprocess the data within one FOV (corresponding to a volume of 14.7 $mm^2 \times 2$ mm) for extracting the MCF7 cell candidates, corresponding to the preliminary screening step in Fig. 6.4. For each detected cell candidate, it takes ~5.5 s to generate the input video for network classification. The network inference time for each input video is <0.01 s. Based on these numbers, if there are, e.g., ~1,500 cell candidates per experiment, the total processing time using the current computer code would be ~3.0 hours. However, we should note that the data processing time depends on various factors, including the computer hardware configuration, the cell concentration in the sample, the programming language and whether the code is optimized for the hardware. In our work, although we used relatively high-performance hardware (an Intel Core i7 CPU, 64 GB of RAM, and an Nvidia GeForce GTX 1080Ti GPU) and used some of the GPU functions provided by MATLAB (MathWorks, Natick, MA, USA), we did not extensively optimize our code for improved speed. A careful optimization of the GPU code should bring a significant speedup in our computation time.

COMSOL simulation of the magnetic force field generated by the electromagnet and the permalloy extension

Because of space constraints, the electromagnet could not be placed sufficiently close to the imaging area, which caused the magnetic force to be low. We used a custom-machined extension rod made of permalloy [215] (relative permeability $\mu_r \sim 100{,}000$) to relay the force field and enhance the relative magnetic force on target cells by ~40 times. To simulate the magnetic force field distribution near an electromagnet with and without the permalloy extension, a finite element method (FEM) simulation was conducted using COMSOL Multiphysics (version 5.3, COMSOL AB, Stockholm, Sweden). A 3D model was developed using the magnetic field interface provided in the COMSOL AC/DC physics package. A stationary study was constructed based on the geometry of a commercially available electromagnet, where the core was modelled with a silicon steel cylinder (radius = 3 mm, height = 10 mm), and the coil was modelled with a surface current of 10 A/m on the side of the core running in the azimuthal direction. The permalloy extension was modelled using Permendur. A thick layer of air was added as a coaxial cylinder with a radius of 10 mm and a height of 30 mm. The magnetic flux density inside the simulation space was simulated using the magnetic field module. Then, a coefficient form PDE module in the mathematics library was used to derive the relative magnetic force field. The magnetic force that is received by superparamagnetic beads is given by:

$$\mathbf{F} = \frac{V\chi}{\mu_0}\left(\mathbf{B}\cdot\nabla\right)\mathbf{B} \tag{61}$$

where V is the volume of the magnetic particle, χ is the magnetic susceptibility, μ_0 is the magnetic permeability in a vacuum, and **B** is the magnetic flux density.

Our simulation results are shown in Fig. S6.1. The results in Fig. S6.1b indicate that the relative magnetic force rapidly reduces as a function of the distance from the electromagnet. However, by using a permalloy extension, the relative magnetic force at the sample location is enhanced by ~40 times.

6.5 Appendix

Table S6.1. Concentrations of different types of cells and particles in the sample before and after the magnetic enrichment. MCF7 cells were spiked into a whole blood sample at a concentration of 1.1×10^5 mL^{-1}, and enrichment was performed following the procedure reported in Fig. 6.2. After the enrichment, the sample was loaded into a counting chamber and imaged by a 20× 0.45NA benchtop microscope, and the particles/cells were manually counted. In order to directly compare it with the concentration before the enrichment step, the concentration after the enrichment is normalized by a volume factor (i.e., the ratio between the volume before the enrichment and the volume after the enrichment). As reported in our table, there is over three orders of magnitude decrease in the concentration of blood cells, while the recovery rate for the target MCF7 cells is ~85%. The concentrations of unbound beads and bead clusters reported here only provide an estimate, because bead clusters are also dynamically formed during the experiments, when there is an external magnetic field applied.

	Concentration (mL^{-1})	
	Before enrichment	After enrichment
Total labeled MCF7 cells	1.1×10^5	9.4×10^4
Non-clustering labeled MCF7 cells	-	4.7×10^4
Labeled MCF7 cell clusters	-	1.7×10^4
Blood cells	$\sim5\times10^9$ (estimated based on the average healthy human blood cell concentration)	1.6×10^6
Magnetic beads	-	1.3×10^6
Bead clusters	-	1.1×10^5

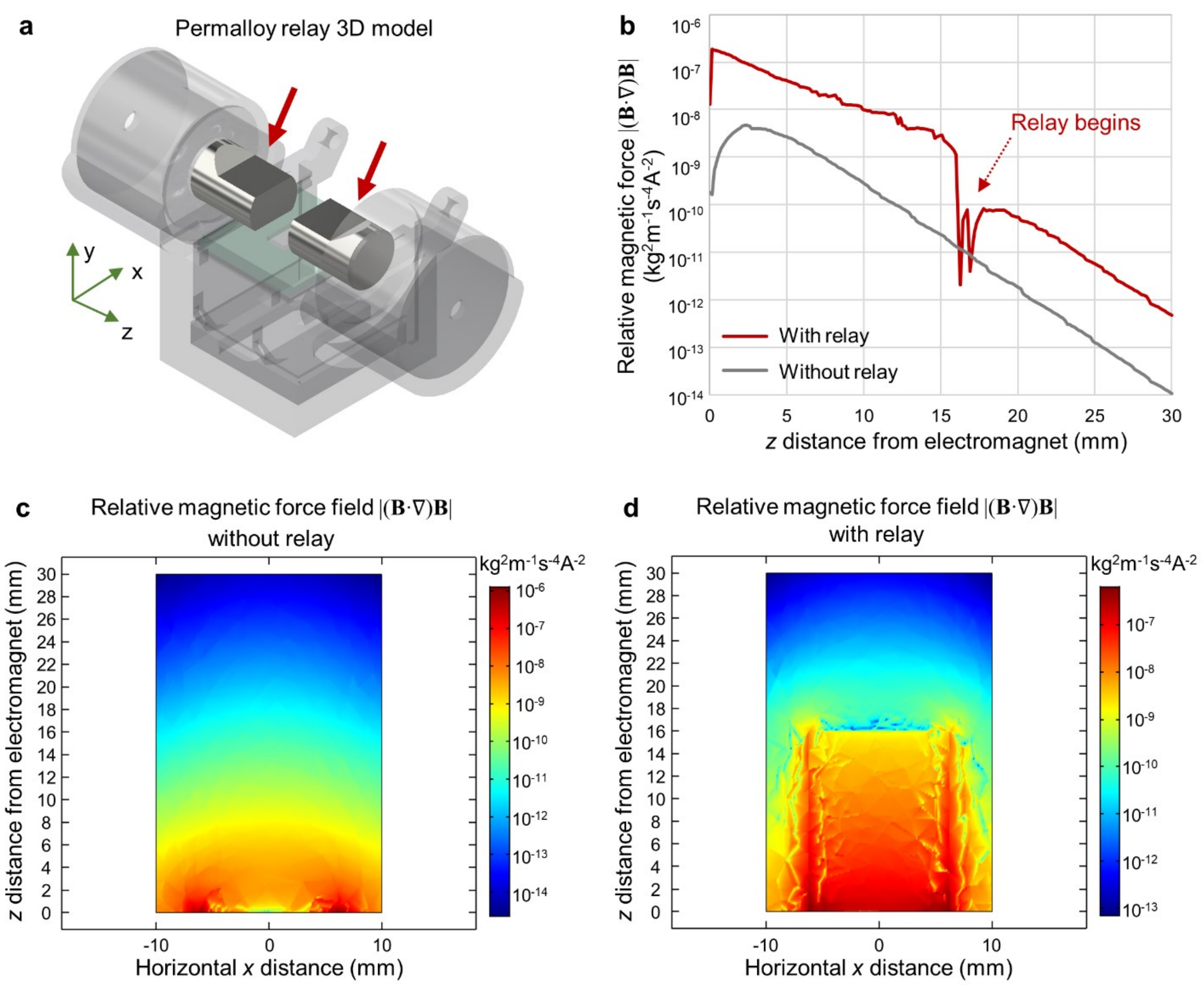


Figure S6.1 COMSOL simulation of magnetic force field generated by electromagnet with permalloy relay. **a** 3D schematic of the permalloy relays relative to the electromagnets. **b-d** Simulation of the relative (unitless) magnitude of the magnetic force field that is generated by a single electromagnet with (**c**) or without (**d**) permalloy relay, as a function of the spatial position. The relay significantly increases the magnetic force field for a given axial distance from the electromagnet.

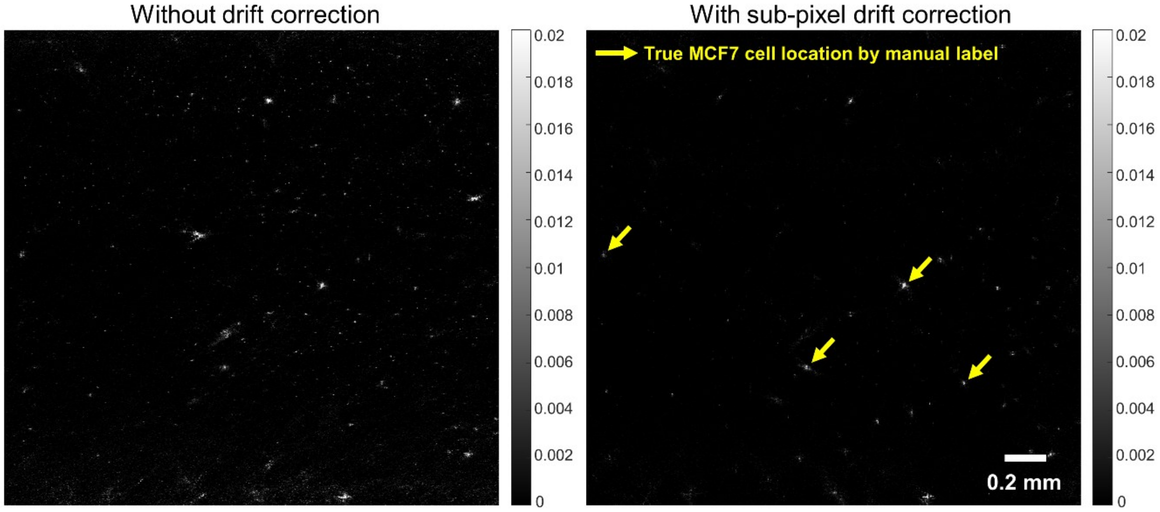


Figure S6.2 Effect of using computational drift correction to remove false positives. Without drift correction, due to the drifting of the medium, particles that do not oscillate in response to the changing magnetic force field may generate contrast in the 2D contrast map, which reduces the effectiveness of the computational motion analysis. The sub-pixel drift correction step removes most of the "false positive" contrast.

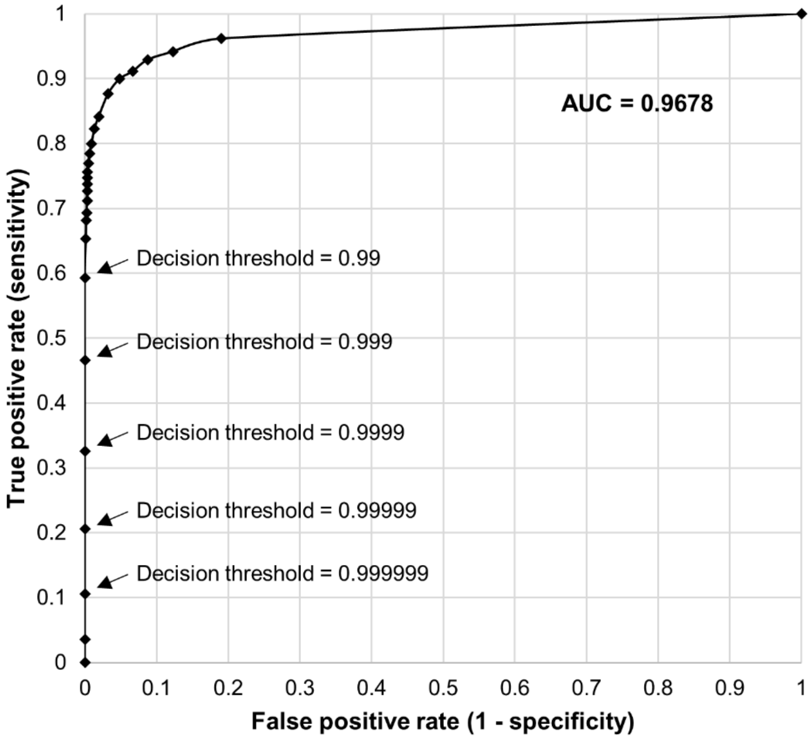


Figure S6.3 Receiver operating characteristic (ROC) curve of the trained P3D CNN classifier. The curve is generated by varying the decision threshold for positive classification using the validation dataset. The area under the curve (AUC) is 0.9678. The blind serial dilution testing experiments (Fig. 6.5) are not used to calculate the AUC, because for those tests the ground truth concentrations are based on the dilution factor used during the spiking experiments. Manual labeling was not performed for the serial dilution testing data.

Chapter 7 . Bacterial colony forming units detection using thin-film-transistor imaging array

7.1 Introduction

Bacterial infection has been a leading factor that causes millions of deaths each year in both developed and developing countries [216,217]. The associated expenses of treating bacterial infections cost more than 4 billion dollars annually in the United States (US) alone [218]. Therefore, the rapid and accurate detection of pathogenic bacteria is of great importance to human health in preventing such infectious diseases caused by e.g., contamination in food and drinking water. Among those pathogenic bacteria, *Escherichia coli* (*E. coli*) and other coliform bacteria are among the most common ones, and they indicate fecal contamination in food and water samples [216]. The most basic and frequently used method of detecting *E. coli* and total coliform bacteria involves culturing the sample on a solid agar plate or liquid medium following the US Environmental Protection Agency (EPA)-approved protocols (e.g., EPA 1103.1 and EPA 1604 methods) [219,220]. However, these traditional culture-based methods usually take ≥24 hours for the final read-out and need visual recognition and counting of colony-forming units (CFUs) by microbiology experts. Although various nucleic acid-based molecular detection approaches [221–224] have been developed for rapid bacteria detection with results ready in less than a few hours, they present lower sensitivity in general and have challenges to differentiate live and dead bacteria [225]; in fact, there is no EPA-approved nucleic acid-based coliform sensing method that can be used for screening water samples. Other bacteria detection methods using different types of biosensors based on e.g., electrochemistry [226,227], fluorescence quenching [228], and surface plasmon resonance (SPR) [229] can also provide rapid results; however, these methods usually need additional surface chemistry or labeling steps; some of these biosensors also provide a positive response even for dead or injured/stressed bacteria, which can create challenges to quantify CFU numbers^{-17}. Various other instruments were developed to provide high sensitivity and specificity for the detection of bacteria based on different methods such as e.g., fluorimetry [233], solid-phase cytometry [234], scatterometry [235], Fresnel diffraction systems [236], fluorescence microscopy [237], Raman spectroscopy [238] and others [239–241]; however, these systems, in general, do not work with large sample volumes (e.g., >0.1 L). As another alternative, Wang *et al.* demonstrated a complementary metal-oxide-semiconductor (CMOS) image sensor-based time-lapse imaging platform to perform early detection and classification of coliform bacteria [225]. This method achieved more than 12 hours of detection time savings and provided species classification with >80% accuracy within 12-hours of incubation. The field-of-view (FOV) of the CMOS image sensor in this design was < 0.3 cm^2, and therefore the mechanical scanning of the Petri dish area was required to obtain an image of the whole FOV of the cultured sample. Not only that this is time-consuming and requires additional sample scanning hardware, but it also brings some extra digital processing burden for image registration and stitching.

Recently, with the fast development of thin-film-transistors (TFT), the TFT technology has been widely used in the field of flexible display industry [242], radio frequency identification tags [243], ultrathin electronics [244], and large-scale sensors [245,246] thanks to its high scalability, low-cost mass production (involving e.g., roll-to-roll manufacturing), low power consumption, and low heat generation properties. TFT technology has also been applied in the biosensing field to detect pathogens by transferring e.g., antibody-antigen binding, enzyme-substrate catalytic activity, or DNA hybridization into electrical signals [246,247]. For example, a low-cost TFT nanoribbon sensor was developed by Hu *et al.* to detect the gene copies of *E. coli* and *Klebsiella pneumoniae (K. pneumoniae)* in a few minutes by using PH change due to DNA amplification [248]. As another example, Salinas *et al.* implemented a ZnO TFT biosensor with recyclable plastic substrates for real-time *E. coli* detection [249,250]. However, these TFT-based biosensing methods could not differentiate between live and dead bacteria and did not provide quantification of the CFU concentration of the sample under test.

Here, we demonstrate the first use of a TFT-based image sensor to build a real-time CFU detection and classification system to automatically count the bacterial colonies and rapidly identify their species using deep learning. To prove the efficacy of our system, *E. coli*, *Citrobacter*, and *K. pneumoniae* were cultured on a growth medium, which provides specificity and color distinction to classify *E. coli* and other total coliforms. Because of the large FOV of the TFT image sensor (~7 cm^2), there is no need for mechanical scanning, which enabled us to create a field-portable and cost-effective lensfree CFU detector as shown in Fig. 7.1. As shown in Fig. 7.1(b), visible light signals are transferred to electrical signals through a detection layer, read through the TFT sensor array [251–253].This compact system includes sequentially switched red, green, and blue light-emitting diodes (LEDs) that periodically illuminate the cultured samples (*E. coli*, *Citrobacter,* and *K. pneumoniae*) as shown in Fig. 7.1(d), and the spatio-temporal patterns of the samples are collected by the TFT image sensor, with an imaging period of 5 min. Two deep learning-based classifiers were trained to detect the bacterial colonies and then classify them into *E. coli* and total coliform bacteria. Blindly tested on a dataset populated with 265 colonies (85 *E. coli* CFU, 66 *Citrobacter* CFU, and 114 *K. pneumoniae* CFU), our TFT-based system was able to detect the presence of the colonies as early as ~6 hours during the incubation period and achieved an average CFU detection rate of 97.3% at 9 hours of incubation, saving more than 12 hours compared to the EPA-approved culture-based CFU detection methods. For the classification of the detected bacterial colonies, an average recovery rate of 91.6% was achieved at ~12 hours of incubation.

This TFT-based field-portable CFU detection system significantly benefits from the cost-effectiveness and ultra-large FOV of TFT image sensors, which can be further scaled up, achieving even lower costs with much larger FOVs based on e.g., roll-to-roll manufacturing methods commonly used in the flexible display industry. We believe that TFT image sensors can potentially be integrated with each agar plate to be tested, and can be disposed of after the determination of the CFU count, opening up various new opportunities for microbiology instrumentation in the laboratory and field settings.

Part of this chapter has been published in:

Li, Y., Liu, T., Koydemir, H.C., Wang, H., O'Riordan, K., Bai, B., Haga, Y., Kobashi, J., Tanaka, H., Tamaru, T., Yamaguchi K., Ozcan, A. (2022) Deep Learning-enabled Detection and Classification of Bacterial Colonies using a Thin Film Transistor (TFT) Image Sensor. ACS Photonics, 9(7), pp.2455–2466

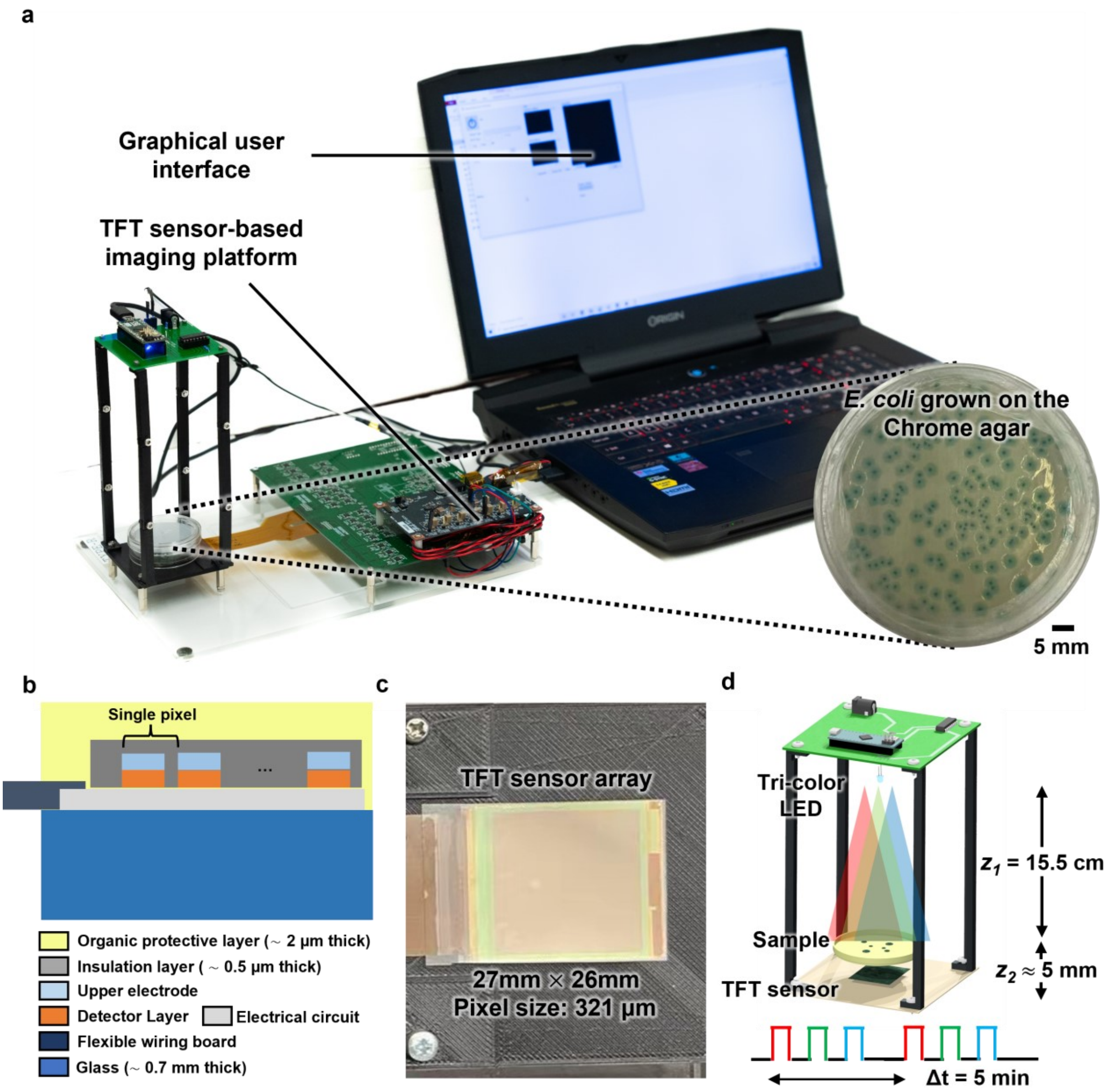


Figure 7.1 Real-time CFU detection and classification system using a TFT image sensor. (a) A photo of the lensfree imaging system, samples to be tested, and the laptop computer used for controlling the hardware. The chromogenic agar medium results in a gray-green color for *E. coli* colonies and a pinkish color for other coliform bacteria; furthermore, it inhibits the growth of different bacterial colonies or exhibits colorless colonies when other types of bacteria are present in the sample. (b) A cross-sectional schematic of the TFT image sensor is shown. (c) A zoomed-in photo of the TFT image sensor with a FOV of ~27 mm × 26 mm. (d) Detailed illustration of the lensfree imaging modality. The red (620 nm), green (520 nm), and blue (460 nm) LEDs were switched on sequentially at 5-minute intervals to directly illuminate the cultured samples, which were imaged by the TFT image sensor in a single shot. The distance between the tri-color LED and the agar plate sample (z_1) is 15.5 cm, while the sample to sensor distance (z_2) distance is ~5 mm.

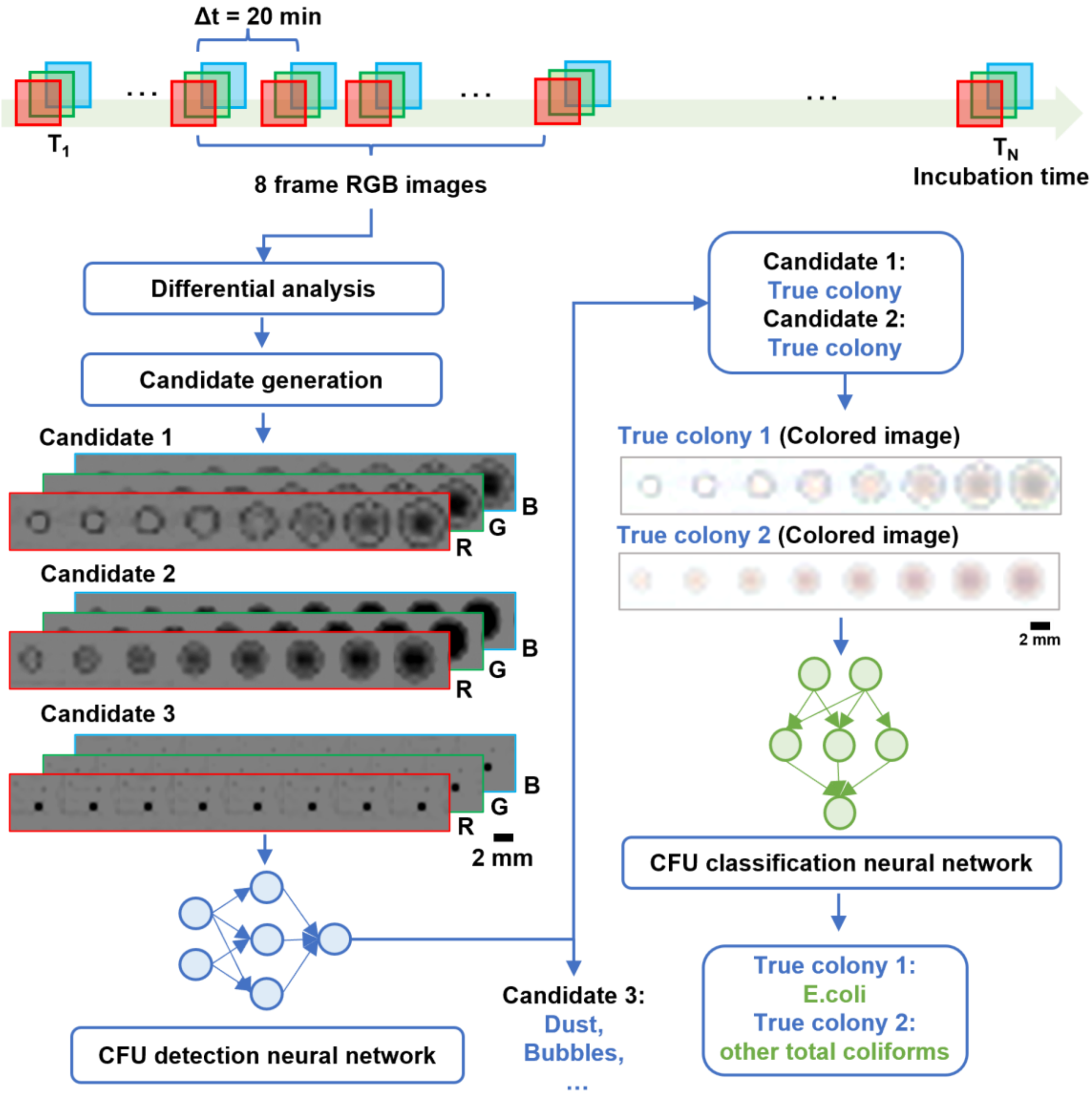


Figure 7.2 Schematics of the workflow of our deep learning-based CFU detection and classification system. We process 8 whole FOV RGB images with 20-minute time intervals for the differential analysis to select the initial colony "candidates". The digitally-cropped 8-frame RGB image sequence for each individual colony candidate is fed into the CFU detection neural network first. This neural network rejects various non-colony objects (among the initial colony candidates) such as dust and bubbles, achieving true colony detection. Next, the detected colonies are passed through the CFU classification neural network to identify their species (*E. coli* or other total coliforms, i.e., binary classification).

7.2 Results

We experimentally demonstrated the success of our framework by detecting and classifying the colonies *E. coli* and two other types of total coliform bacteria, i.e., *Citrobacter* and *K. pneumoniae,* on chromogenic agar plates, which result in a gray-green color for *E. coli* colonies and a pinkish color for other coliform bacteria, also inhibiting the growth of different bacterial colonies when other types of bacteria exist in the sample. Each sample was prepared following the EPA-1103.1 method [220] (see the Methods section) using a Petri dish with 60-mm diameter. After the sample was prepared, it was directly placed on top of the TFT image sensor as part of the lensfree imaging system, which images the center of the sample over a FOV of ~27 mm × 26 mm, missing the edges of the Petri dish. This FOV can be further increased to cost-effectively cover the whole area of a Petri dish using industrial roll-to-roll TFT manufacturing methods. The entire imaging modality (except the laptop in Fig. 7.1(a)) was placed inside an incubator to record the growth of the colonies with 5-minute imaging intervals using the automated TFT readout. Each sample was left in the incubator for 24 hours to let the colonies further grow for manual readout, which served as the ground truth measurement. For each time interval, three images were collected sequentially using the TFT image sensor under red (620 nm), green (520 nm), and blue (460 nm) illumination light. This multi-wavelength design allowed the monochromatic TFT image sensor to reconstruct color images of the bacterial colonies and was mainly used to identify their species by exploiting the color information provided by the selective chromogenic agar medium. The recorded time-lapse images were processed using the workflow shown in Fig. 7.2, where a differential analysis was used to select the initial colony candidates, and two deep neural networks (DNNs) were trained to further screen the colony candidates to specifically detect the true colonies and infer their species (see the Methods section for details). All these image processing steps take <25 sec using an Intel Core i7-7700 CPU-powered computer, consuming <1 GB of memory (without the need for GPUs).

The presented TFT imaging system periodically captures the images of the agar plate under test based on lensfree in-line holography; however, due to its large pixel size (321 μm) and relatively small sample to sensor distance (~5 mm, which is equal to the thickness of the agar), a free space backpropagation [254–258] step is not needed. By directly using the raw intensity images as part of the RGB color channels and calibrating the background, the color images of the agar plate can be generated in <0.25 sec after the TFT images are recorded. Fig. 7.3 shows examples of color images of *E. coli*, *Citrobacter,* and *K. pneumoniae* colonies at different stages of their growth, captured by our system. Consistent with the EPA-approved method (EPA-1103.1 [220]), *E. coli* colonies exhibit gray-green colors, while *Citrobacter* and *K. pneumoniae* colonies exhibit pinkish color using the chromogenic agar.

Based on the imaging performance of our TFT-based CFU detection system summarized in Fig. 7.3, we quantified its early detection and classification performance as shown in Fig. 7.4. For this, we trained the detection and the classification neural network models (see the Methods section for training details) on a dataset of 442 colonies (128 *E. coli* colonies, 126 *Citrobacter,* and 188 *K. pneumoniae* colonies) captured from 17 independent experiments. The testing dataset was populated using 265 colonies from 13 independent experiments, which had a total of 85 *E. coli* colonies, 66 *Citrobacter* colonies, and 114 *K. pneumoniae* colonies. The detection rate was defined as the ratio of the number of true colonies confirmed by the CFU detection neural network out of the total colony number counted by an expert after 24-hour incubation. Fig. 7.4(a, c, e) shows the detection rate we achieved in the blind testing phase as a function of the incubation time. As shown in Fig. 7.4(a, c, e), > 90% detection rate was achieved at 8 hours of incubation for *E. coli*, 9 hours for *Citrobacter*, and 7 hours 40 minutes for *K. pneumoniae*. Furthermore, a 100% detection rate was obtained within 10 hours of incubation for *E. coli*, 11 hours for *Citrobacter*, and 9 hours 20 minutes for *K. pneumoniae*. Compared to the EPA-approved standard read-out time (24 hours), our TFT-based CFU detection system achieved > 12 hours of time-saving. Moreover, from the detection rate curves reported in Fig. 7.4, we can also qualitatively infer that the colony growth speed of *K. pneumoniae* is larger than *E. coli* which is larger than *Citrobacter* because the earliest detection times for *E. coli*, *Citrobacter,* and *K. pneumoniae* colonies were 6 hours, ~6.5 hours and ~5.5 hours of incubation, respectively.

To quantify the performance of our bacterial colony classification neural network, the recovery rate was defined as the ratio of the number of correctly classified colonies to the total number of colonies counted by an expert after 24-hour incubation. Fig. 7.4(b, d, f) shows the recovery rate curves over all the blind testing experiments as a function of the incubation time. We can see that a recovery rate of > 85% was achieved at 11 hours 20 minutes for *E. coli*, at 13 hours for *Citrobacter*, and at 10 hours 20 minutes for *K. pneumoniae.* It is hard to achieve a 100% recovery rate for all the colonies since some of the late growing "wake-up" colonies could not grow to a sufficiently large size with the correct color information even after 24 hours of incubation. Fig. 7.4 also reveals that there exists approximately a 3-hour time delay between the colony detection time and species identification time; this time delay is expected since more time is needed for the detected colonies to grow larger and provide discernable color information for the correct classification of their species.

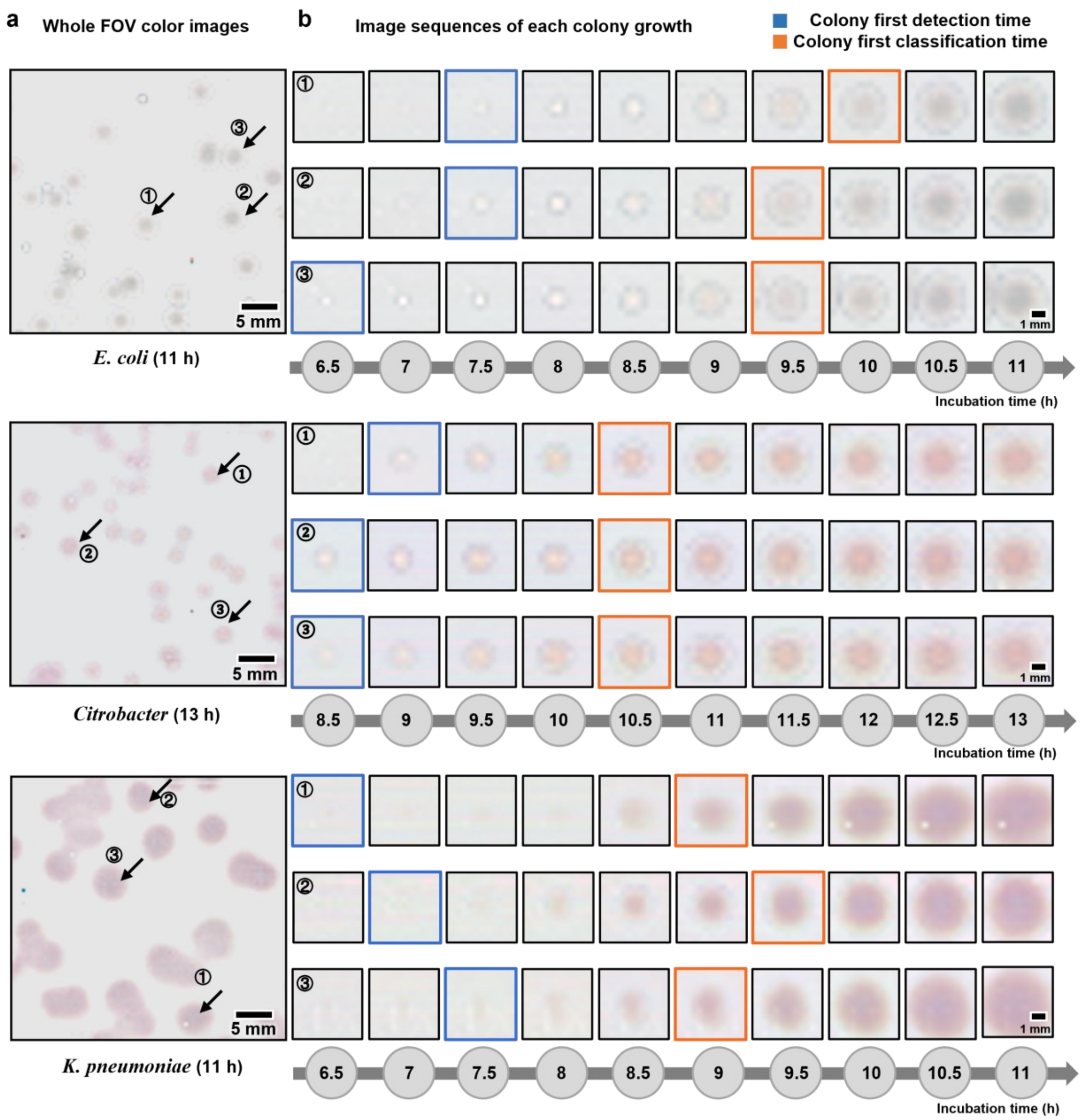


Figure 7.3 Visual evaluation of coliform bacterial colony early detection and classification using a TFT image sensor. (a) Whole FOV color images of *E. coli* at 11-hour incubation, *Citrobacter* at 13-hour incubation, and *K. pneumoniae* at 11-hour incubation. (b) Examples of the image sequence of each isolated colony growth. Three independent colony growth sequences were selected for each one of the bacteria species. The blue box labels the first colony detection time confirmed by the CFU detection neural network, and the orange box corresponds to the first classification time correctly predicted by the CFU classification neural network.

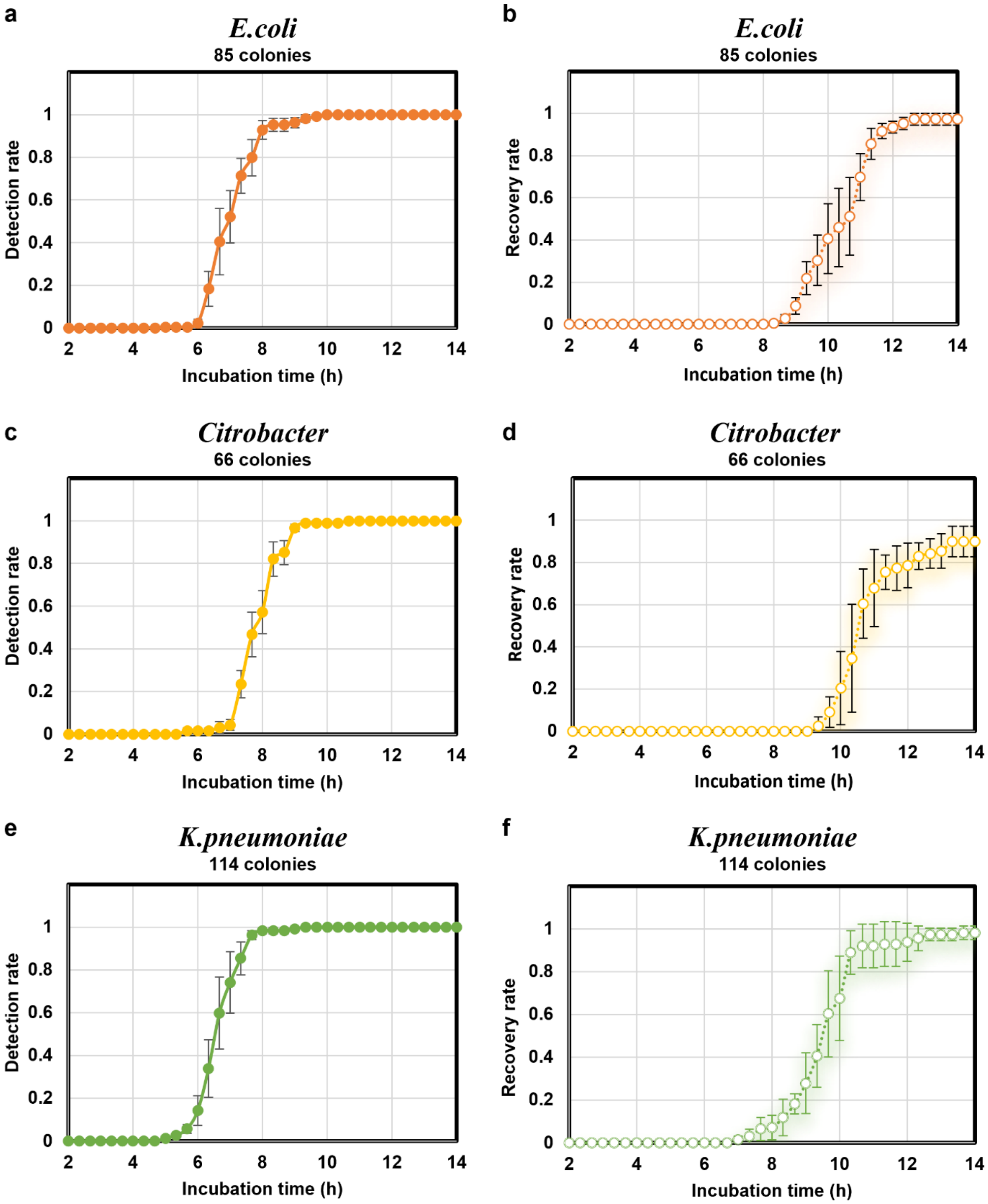


Figure 7.4 Quantitative performance evaluation of coliform colony early detection and classification using a TFT image sensor. (a, c, e) The colony detection rate as a function of the incubation time for *E. coli*, *Citrobacter,* and *K. pneumoniae*. The mean and standard deviation of the detection rate were calculated on 85 *E. coli* colonies, 66 *Citrobacter* colonies, and 114 *K. pneumoniae* colonies for each time point. (b, d, f) The colony recovery rate as a function of the incubation time for *E. coli*, *Citrobacter,* and *K. pneumoniae*. The mean and standard deviation of the recovery rate were calculated on 85 *E. coli* colonies, 66 *Citrobacter* colonies, and 114 *K. pneumoniae*

colonies for each time point.

7.3 Discussion

Note that our presented results in Fig. 7.4 represent a conservative performance of our TFT-based CFU detection method since the ground truth colony information was obtained after 24 hours of incubation. In the early stages of the incubation period, some bacterial colonies did not even exist physically. Therefore, if we used the existing colony numbers for each time point as the ground truth, we would have reported even higher detection and recovery rates in Fig. 7.4.

Overall, the performance of our TFT-based CFU detection method is similar to the CMOS-based time-lapse imaging method [225] in terms of the colony detection speed. However, due to its large pixel size (321 μm) and limited spatial resolution, the TFT-based method has a slightly delayed colony classification time. With its ultra-large imaging FOV (~7 cm^2), the TFT-based CFU detection method eliminates (1) the time-consuming mechanical scanning of the Petri dish and the related optomechanical hardware, and (2) the image processing steps for image registration and stitching that would both be required due to the limited FOV of CMOS-based imagers. In addition to saving image processing time, this also helps the system to increase the CFU detection sensitivity as the system is free from any image registration and stitching artifacts and therefore, it can precisely capture minute spatio-temporal changes in the agar caused by bacterial colony growth at an early stage. Due to the massive scalability of the TFT arrays, the imaging FOV of our platform can be further increased to several tens to hundreds of cm^2 in a cost-effective manner, which could provide unprecedented levels of imaging throughput for automated CFU detection using e.g., roll-to-roll manufacturing of TFTs, as employed in the flexible display industry.

Another prominent advantage of the TFT-imager based detection system is that it can be adapted to image a wide range of biological samples using cost-effective and field-portable interfaces. As shown in the blue circled regions in Fig. 7.3, our system can detect bacterial colonies in their early stage of growth with colorless and low contrast features. By using new training data and deep learning, our method can be adapted to specifically and sensitively detect and classify other bacterial colonies using different growth media. Furthermore, should the users have any contamination concerns, the TFT image sensor shown in Fig. 7.1 can be replaced and even used in a disposable manner (e.g., integrated as part of the Petri dish). Furthermore, the heat generated by the TFT image sensor during the data acquisition process is negligible, ensuring that the biological samples can grow at their desired temperature without being perturbed. Finally, our TFT-based CFU detection system is user-friendly and easy-to-use because there is no need for complex optical alignment, high precision mechanical scanning stages, or image registration/alignment steps.

In summary, we believe that the presented CFU detection system using TFT image sensor arrays provides a high-throughput, cost-effective, and easy-to-use solution to perform early detection and classification of bacterial colonies, opening up unique opportunities for microbiology instrumentation in the laboratory and field settings.

7.4 Materials and methods

Sample preparation

All the bacterial sample preparations were performed at our Biosafety Level 2 laboratory in accordance with the environmental, health, and safety rules of the University of California, Los Angeles. We used *E. coli* (Migula) Castellani and Chalmers (ATCC® 25922™), *Citrobacter* (ATCC® 43864™), and *K. pneumoniae* subsp. *pneumoniae* (Schroeter) Trevisan (ATCC®13883™) as our culture microorganisms. CHROMagar™ ECC (product no. EF322, DRG International, Inc., Springfield, NJ, USA) chromogenic substrate mixture was used as the solid growth medium to detect *E. coli* and other total coliform colonies.

For each time-lapse imaging experiment, a bacterial suspension in a phosphate-buffered solution (PBS) (product no. 20-012-027, Fisher Scientific, Hampton, NH, USA) was prepared from a solid agar plate incubated for 24 hours. The concentration of the suspension was measured using a spectrophotometer (model no. ND-ONE-W, Thermo Fisher). Then, a serial dilution was performed in PBS to finally reach a concentration of ~10^3 CFUs / mL. Around 100 μL diluted suspension with ~100 CFUs was spread on a CHROMagar™ ECC plate using an L-shaped spreader (product no. 14-665-230, Fisher Scientific, Hampton, NH, USA). Next, the plate was covered with its lid, inverted, and placed on the TFT image sensor, which was placed with the whole imaging system into an incubator (product no. 151030513, ThermoFisher Scientific, Waltham, MA, USA) kept at 37 ± 0.2 °C.

Additionally, CHROMagar™ ECC plates were prepared ahead of time using the following method. CHROMagar™ ECC (6.56 g) was mixed with 200 mL of reagent grade water (product no. 23-249-581, Fisher Scientific, Hampton, NH, USA). The mixture was then heated to 100 °C on a hot plate while being stirred regularly using a magnetic stirrer bar. After cooling the mixture to ~50 °C, 10 mL of the mixture was dispensed into each Petri dish (60 mm × 15 mm) (product no. FB0875713A, Fisher Scientific, Hampton, NH, USA). When the agar plates solidified, they were sealed using parafilm (product no. 13-374-16, Fisher Scientific, Hampton, NH, USA), and covered with aluminum foil to keep them in the dark before use. These plates were stored at 4 °C and were used within two weeks after preparation.

Imaging Set-up

Our field-portable CFU imager comprises an illumination module and a TFT-based image sensor. The light from a tri-color LED directly illuminates the samples and forms in-line holograms on the TFT image sensor (JDI, Japan Display Inc., Japan). The TFT module includes a controlling printed circuit board (PCB) that provides the illumination and image capture control signal and an image sensor (with 84×80 pixels, pixel size = 321 μm). For the illumination module, a tri-color LED (EDGELEC) was controlled by a microcontroller (Arduino Micro, Arduino LLC) through a constant current LED driver (TLC5916, Texas Instrument, TX, USA) to sequentially provide the red (620 nm), green (520 nm), and blue (420 nm) illumination beams. The microcontroller, the LED driver, and the tri-color LED were all integrated on a single PCB, which was powered by a 5V-1A voltage adapter and communicated with the TFT PCB through the LED power signal.

The illumination light passes through the transparent solid agar and forms the lensfree images of the growing bacterial colonies on the TFT image sensor. The distance between the LED and the sample (i.e., the z_1 distance shown in Fig. 7.1(c)), is ~15.5 cm, which is large enough to make the illumination light uniformly cover the whole sample surface. The distance between the sample and the sensor (z_2) is roughly equal to the thickness of the solid agar, which is ~5 mm. The mechanical support material for the PCB, the sample, and the sensor were custom fabricated using a 3D printer (Objet30 Pro, Stratasys, Minnesota, USA).

Image data acquisition

Time-lapse imaging experiments were conducted to collect the data for both the training and testing phases. Our CFU imaging modality captured the time-lapse images of the agar plate under test every 5 min under red, green, and blue illuminations. A controlling program with a graphical user interface (GUI) was developed to perform the illumination switching and image capture automatically. The raw TFT images were saved in 12-bit format. After the experiments were completed, the samples were disposed of as solid biohazardous waste. In total, we collected the time-lapse TFT images of 889 *E. coli* colonies from 17 independent experiments to initially train our CFU detection neural network model. In addition to this, 442 bacterial colonies (128 *E. coli*, 126 *Citrobacter*, and 188 *K. pneumoniae*) were populated from 17 new agar plates and used to train (1) the final CFU detection neural network (through transfer learning from the initial detection model) and (2) the CFU classification neural network. A third independent dataset of 265 colonies from 13 new experiments was used to test the trained neural network models blindly.

Bacterial colony candidate selection

The entire candidate selection workflow consists of image pre-processing, differential analysis, colony mask segmentation, and candidate position localization, following the steps listed in Fig. 7.5(a-i). For each time point, three raw TFT images (red, green, and blue channels) were obtained over a FOV of ~7 cm^2. After getting the TFT images $I_{N\_\mathrm{raw},\,C}$, where N refers to the N-th image obtained at T_N and C represents the color channels, R (red), G (green), and B (blue), a series of pre-processing operations were performed to enhance the image contrast. First, as shown in Fig. 7.5(a-b), the images were 5 times interpolated and normalized by directly subtracting the first frame at T_0. After this normalization step, the background regions had ~0 signal, while the regions representing the growing colonies had negative values because the colonies partially blocked and scattered the illumination light. Then, by adding 127 and saving the images as unsigned 8-bit integer arrays, the current frame at T_N was scaled to 0-127, noted as $I_{N\_\mathrm{norm},\,C}$.

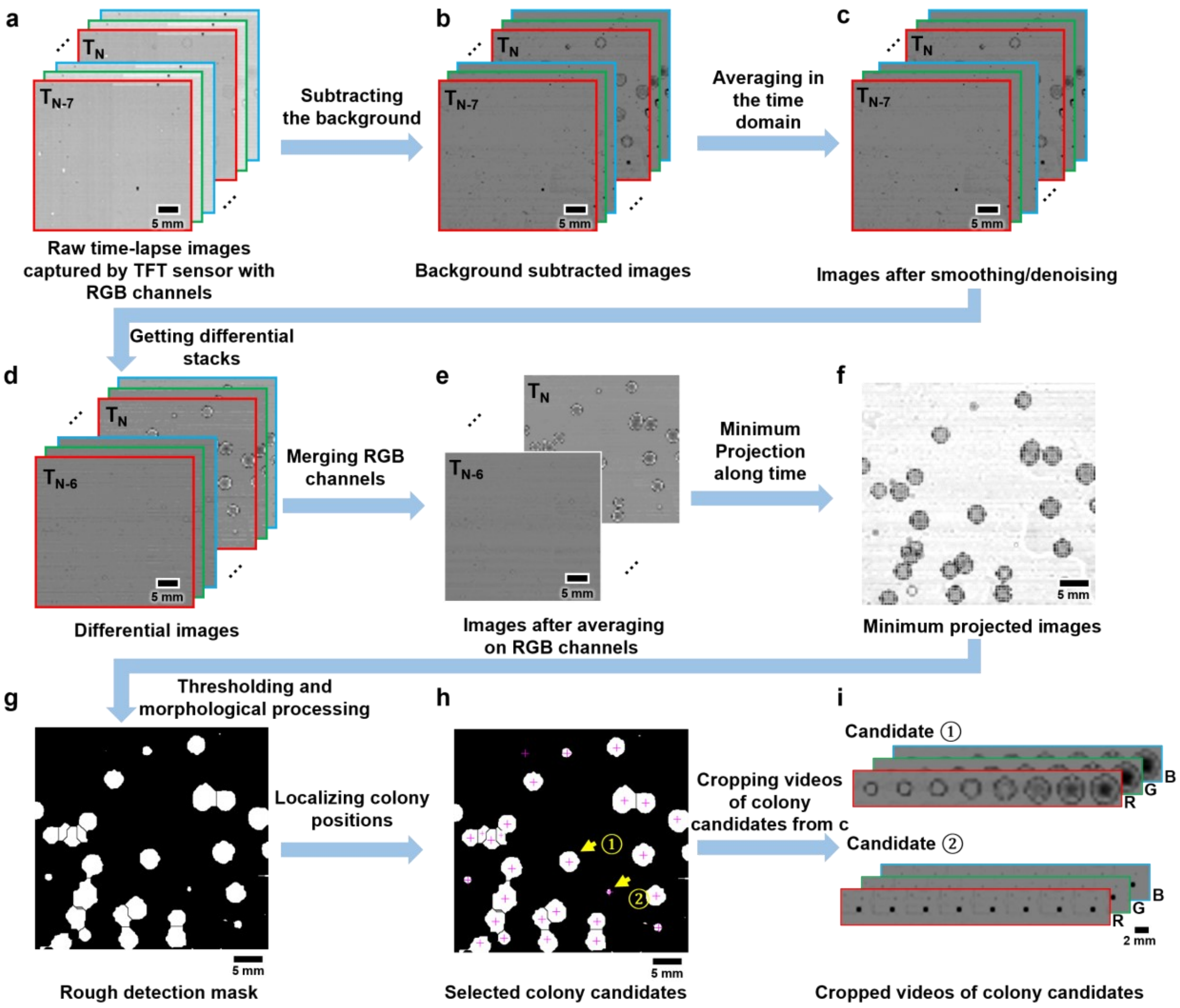


Figure 7.5 Bacterial colony candidate generation workflow. The image pre-processing steps (a-i) were performed on the acquired TFT images in order to select the colony candidates; the cropped videos of the colony candidates were then passed through a trained CFU detection neural network to determine the true positives and eliminate false positives.

Following the steps in Fig. 7.5 (b-c), $I_{N\_\mathrm{norm},\,C}$ was averaged as shown in Equation 62 to perform smoothing in the time domain, which yields $I_{N\_\mathrm{denoised},\,C}$:

$$I_{N\_\mathrm{denoised},C} = \frac{1}{3}\sum_{n=N-2}^{n=N} I_{n\_\mathrm{norm},C} \tag{62}$$

To further improve the sensitivity of our system, differential images I_{N_diff} averaged on three color channels were calculated as follows:

$$I_{N\_\mathrm{diff}} = \frac{1}{3}\sum_{C=\mathrm{R,G,B}}\left(I_{N\_\mathrm{denoised},C} - I_{(N-1)\_\mathrm{denoised},C}\right) \tag{63}$$

By this operation, the signals of static artifacts were suppressed, and the spatio-temporal signals of the growing colonies were enhanced as ring-shaped patterns. Next, a pixel-wise minimum intensity projection was performed, as shown in Fig. 7.5(e-f), to project the minimum intensity of the differential images from $I_{(N\text{-}7)\_\text{diff}}$ to I_{N_diff}, yielding the image $I_{N\_\text{projection}}$. Following this step, with an empirically set intensity threshold, $I_{N\_\text{projection}}$ was segmented into a binary mask. After morphological operations to fill the ring-shaped patterns and a watershed-based [259] division of clustered regions, M_N was obtained as presented in Fig. 7.5(g). Based on this binary mask, M_N, we extracted the connected components and localized their centroids as shown in Fig. 7.5(h). These centroid coordinates were dynamically updated for each time point to ensure maintaining the localization at the center of the growing colonies.

Despite this pre-processing of the acquired TFT images, there are still some time-varying non-colony objects that can be selected as false colony candidates (such as bubbles, dust, or other features created by the uncontrolled motion of the agar surface). Therefore, a deep neural network was trained to further screen each colony candidate to eliminate false positives, the details of which will be discussed in the next subsection.

DNN-based detection of bacterial colony growth

The time-lapse video of each colony candidate region across 8 frames of $I_{N\_\text{denoised},\ C}$ was cropped as shown in Fig. 7.5. These videos were then up-sampled in the spatial domain and organized as a four-dimensional array (3×8×160×160, i.e., color channels×number of frames ×x×y) to be fed into the CFU detection neural network, which adopted the architecture of Dense-Net [260], but with 2D convolutional layers replaced by pseudo-3D convolutional layers [261] (see Fig. 7.6). The weights of this CFU detection DNN were initialized with a pre-trained model obtained on the *E. coli* CFU dataset with a single illumination wavelength of 515 nm. This pre-trained model was obtained using a total of 889 colonies (positives) and 159 non-colony objects (negatives) from 17 independent agar plates. Then, this initial neural network model was transferred to the multiple-species image dataset with multi-wavelength illumination, using 442 new colonies and 135 non-colony objects from another 17 independent agar plates. Both the positive image dataset and the negative image dataset were augmented across the time domain with different starting and ending time points, resulting in more than 10,000 videos used for training. A 5-fold cross-validation strategy was adopted to select the best hyper-parameter combinations. Once the hyper-parameters were decided, all the collected data were used for training to finalize our CFU detection neural network model. Data augmentation, such as flipping and rotation, was also applied when loading the training dataset.

The network model was optimized using the Adam optimizer with a momentum coefficient of (0.9, 0.999). The learning rate started as 1×10^{-4} and a scheduler was used to decrease the learning rate with a coefficient of 0.8 at every 10 epochs. The batch size was set to 8. The loss function was selected as:

$$l(p,g)=\sum_{k=1}^{k=K}\sum_{l=1}^{2}-w_l\cdot g_{k.l}\cdot\log\left(\frac{\exp(p_{k,l})}{\exp(p_{k,1})+\exp(p_{k,2})}\right) \tag{64}$$

where p is the network output, which is the probability of each class before the SoftMax layer, g is the ground-truth label (which is equal to 0 or 1 for binary classification), K is the total number of training samples in one batch, w is the weight assigned to each class, defined as $w=1-d$ where d is the percentage of the samples in one class. The training process was performed using a GPU (GTX1080Ti) which took ~5 hours to converge. With a decision threshold of 0.5, the CFU detection neural network converged with 92.6% sensitivity and 95.8% specificity. In the testing phase, the decision threshold was set to be 0.99, which achieved 100% specificity.

DNN-based classification of *E. coli* and other total coliform colonies

To classify the species of the detected bacterial colonies, a second DNN-based classifier was built. The CFU classification neural network was trained on the same multi-wavelength dataset populated with 442 colonies (128 *E. coli* colonies, 126 *Citrobacter* colonies, and 188 *K. pneumonia* colonies). The input of the classification DNN was organized into a four-dimensional array (3×8×160×160, i.e., color channels×number of frames×x×y), but with a different normalization method. Different from the background subtraction normalization adopted for the CFU detection neural network, for the classification DNN, the network input was re-normalized by dividing the background intensities obtained at the first time point T_0. This division-based normalization was performed on three color channels so that the background would be normalized to ~1 in the three channels, revealing a white color in the background. Through this operation, the color variations across different experiments were minimized, improving the generalization capability of the classification DNN.

The network structure of the classification DNN was the same as the CFU detection network but with some differences in the hyper-parameter selection (see Fig. 7.6). The classification neural network model was initialized randomly and optimized using the Adam optimizer with a momentum coefficient of (0.9, 0.999). The learning rate started with 1×10^{-3} and a scheduler was used to decrease the learning rate with a coefficient of 0.7 at every 30 epochs. The batch size was also set to 8. The classification neural network also used the weighted cross-entropy loss function as shown in Equation 64. The training process was performed using a GPU (GTX1080Ti) which took ~5 hours to converge. A decision threshold of 0.5 was used to classify the E. coli colonies and other total coliform colonies in the training process, achieving 91% and 97% accuracy, respectively. In the testing phase, the decision threshold was set to be 0.8, which achieved 100% classification accuracy. In addition, a colony size threshold of 4.5 mm^2 was used in the testing phase to ensure that only colonies that are large enough to identify their species were passed through the classification network.

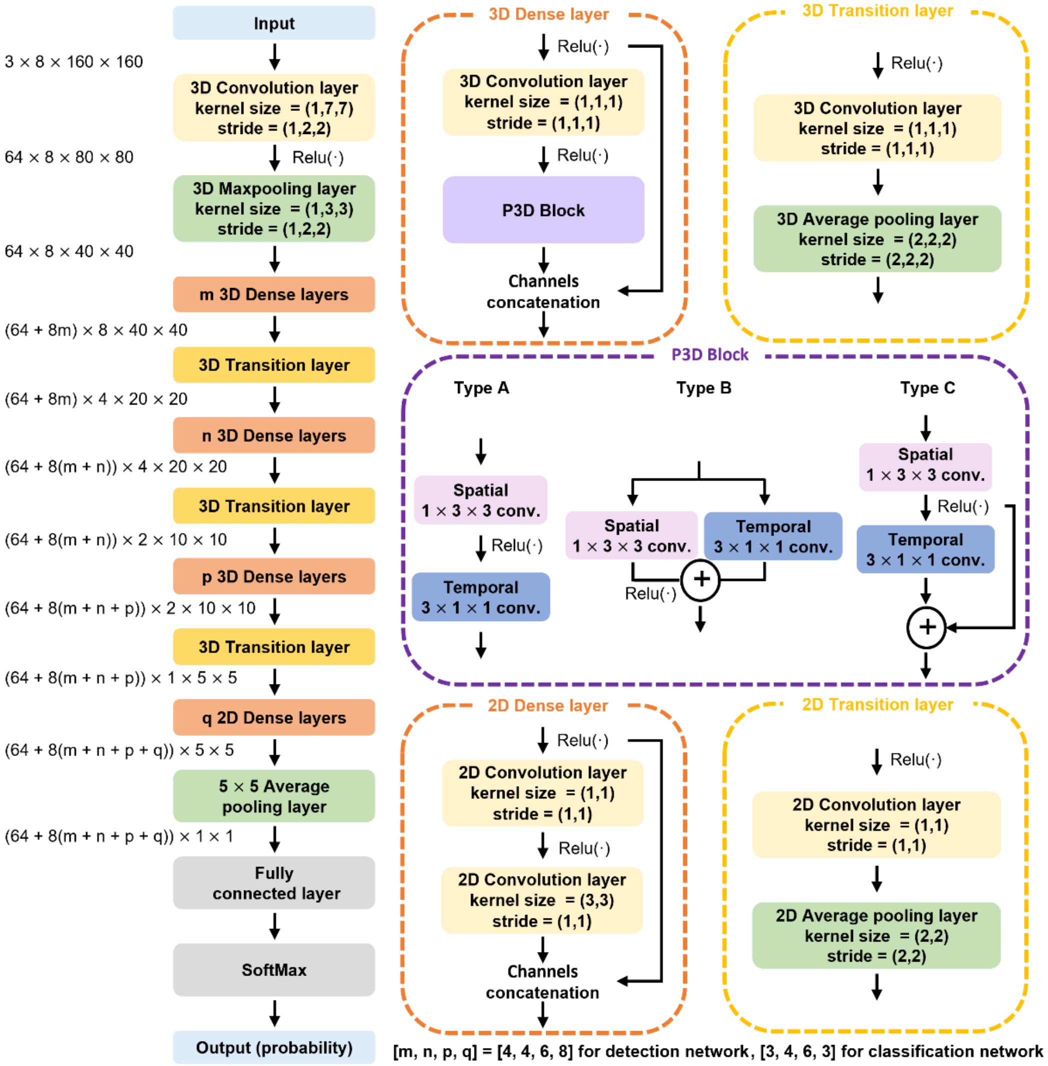


Figure 7.6 Network architectures for the CFU detection neural network and the CFU classification neural network. A Dense-Net design was adopted here, with the 2D convolutional layers replaced by the pseudo-3D convolutional blocks. The CFU detection and classification neural networks shared the same architecture, but the hyper-parameters [m, n, p, q] are selected to be different.

Chapter 8 . Stain-free viral plaque assay using deep learning and holography

8.1 Introduction

Viral infections pose significant global health challenges by affecting millions of people worldwide through infectious diseases, such as influenza, human immunodeficiency virus (HIV), human papillomavirus (HPV), and others [262]. The US Centers for Disease Control and Prevention (CDC) estimates that, since 2010, the influenza virus has resulted in 16-53 million illnesses, 0.2-1 million hospitalizations, and 16,700-66,000 deaths in the United States alone [263,264]. Furthermore, the ongoing COVID-19 pandemic has already caused >500 million infections and >6 million deaths worldwide, bringing a huge burden on public health and socioeconomic development [265]. To cope with these global health challenges, developing an accurate and low-cost virus quantification technique is crucial to clinical diagnosis [266], vaccine development [267], and the production of recombinant proteins [268] or antiviral agents [269,270].

Plaque assay was developed as the first method for quantifying virus concentrations in 1952 and was advanced by Renato Dulbecco, where the number of plaque-forming units (PFUs) was manually determined in a given sample containing replication-competent lytic virions [271,272]. These samples are serially diluted, and aliquots of each dilution are added to a dish of cultured cells [271]. As the virus infects adjacent cells and spreads, a plaque will gradually form, which can be visually inspected by an expert. Due to its unique capability of providing the *infectivity* of the viral samples in a *cost-effective* way, the plaque assay remains to be the gold standard method for quantifying virus concentrations despite the presence of other methods [273–280] such as the immunofluorescence focal forming assays (FFA) [275], polymerase chain reaction (PCR) [277], and enzyme-linked immunoassay (ELISA) based assays [280,281]. However, plaque assays usually need an incubation period of 2-14 days (depending on the type of virus and culture conditions) [282] to let the plaques expand to visible sizes, and are subject to human errors during the manual plaque counting process [283]. To improve the traditional plaque assays, numerous methods have been developed [284]. While these earlier systems have unique capabilities to image cell cultures in well plates, they require either fluorescence markers [283] or special culture plates with gold microelectrodes [285]. In addition, human counting errors still remain to be a problem for these methods. Hence, an accurate, quantitative, automated, rapid, and cost-effective plaque assay is urgently needed in virology research and related clinical applications.

Some of the recent developments in quantitative phase imaging (QPI), holography, and deep learning provide an opportunity to address this need. QPI is a preeminent imaging technique that enables the visualization and quantification of transparent biological specimens in a non-invasive and label-free manner [286,287]. Furthermore, the image quality of QPI systems can be enhanced using neural networks by improving e.g., phase retrieval [288], noise reduction [130], auto-focusing [289,290], and spatial resolution [291]. In addition, numerous deep learning-based microorganism detection and identification methods have been successfully demonstrated using QPI [225,292–296,138,297–300].

Here, we report a cost-effective and compact label-free live plaque assay that can automatically provide significantly faster quantitative PFU readout than traditional viral plaque assays without the need for staining. A compact lensfree holographic imaging prototype was built (Fig. 8.1) to image the spatio-temporal features of the target PFUs during their incubation; the total cost of the parts of this entire imaging system is < \$880, excluding a standard laptop computer. This lensfree holographic imaging system rapidly scans the entire area of a 6-well plate every hour (at a throughput of ~0.32 Giga-pixels per scan of a test well), and the reconstructed phase images of the sample are used for PFU detection based on the spatio-temporal changes observed within the wells. A neural network-based classifier was trained and used to convert the reconstructed phase images to PFU probability maps, which were then used to reveal the locations and sizes of the PFUs within the well plate. To prove the efficacy of our system, early detection of vesicular stomatitis virus (VSV) was performed on Vero E6 cell plates. Our stain-free device could automatically detect the first cell-lysing event due to the virus replication as early as 5 hours after the incubation and achieve >90% PFU detection rate in <20 hours, providing major time savings compared to the traditional plaque assays that take ≥48 hours. A quantitative relationship was also developed between the incubated virus concentration and the virus-infected area on the cell monolayer. Without any extra sample preparation steps, this deep learning-enabled label-free PFU imaging and quantification device can be used with various plaque assays in virology and might help to expedite vaccine and drug development research.

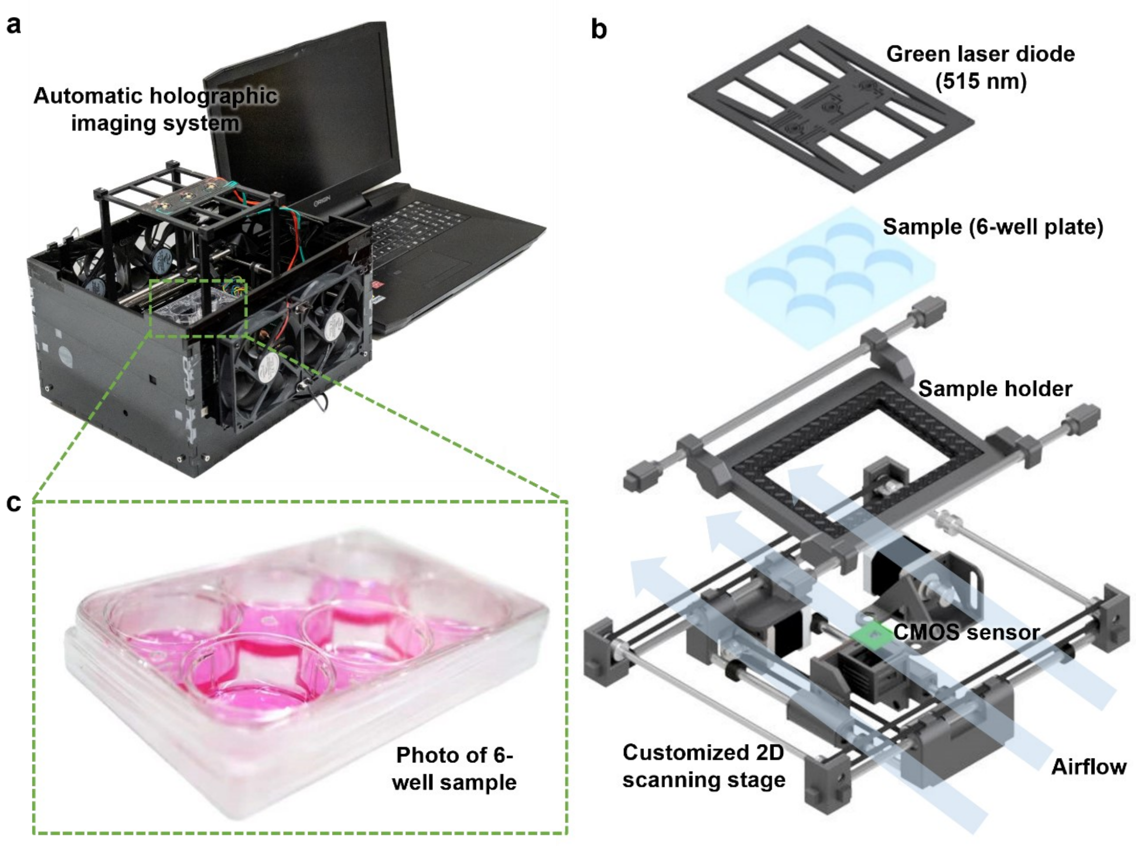


Figure 8.1 Stain-free, rapid and quantitative viral plaque assay using deep learning and lensless holography. (a) Photograph of the stain-free PFU imaging system that captures the phase images of the plaque assay at a throughput of ~0.32 Giga-pixels per scan of each test well. The processing of each test well using the PFU classifier network takes ~7.5 min/well, automatically converting the holographic phase images of the well into a PFU probability map (see Fig. 8.2). (b) Detailed illustration of the system components. (c) A 6-well plate sample with ventilation holes on the cover and parafilm sealed from the side.

Part of this chapter has been published in:

Liu, T., Li, Y., Koydemir, H.C., Zhang, Y., Yang, E., Wang, H., Li, J., Bai, B. and Ozcan, A., (2022). Stain-free, rapid, and quantitative viral plaque assay using deep learning and holography. arXiv preprint arXiv:2207.00089.

8.2 Results

To demonstrate the efficacy of the presented device, we prepared 14 plaque assays using the Vero E6 cells and VSV. The sample preparation steps are summarized in Fig. 8.2a (also see the Methods section); for each well-plate, 5 wells were infected by the VSV and 1 well was left for negative control. After each sample was prepared, it was first placed into our imaging set-up for 20 hours of incubation, performing time-lapse imaging to capture the spatio-temporal information of the sample. Then, the same sample was left in the incubator for an additional 28 hours to let the PFUs grow to their optimal size for the traditional plaque assay (this is only used for comparison purposes). Finally, each sample was stained using crystal violet solution to serve as the ground truth to compare against our label-free method.

To train and test our network-based PFU classifier, 54 wells were used for training and 30 wells were used for testing. During the training phase, a machine learning-based coarse PFU localization algorithm was developed to both accelerate the training dataset generation and delineate the potential false positives (see the Method sections for details). After this PFU localization algorithm screened each sample, the resulting PFU candidates were further examined manually for confirmation using a custom-developed Graphical User Interface (Supplementary Figure 8.1); this manual examination was only used during the training phase to correctly and efficiently prepare the training data. The negative training dataset was populated purely from the negative control well of each well plate. In total, 357 true positive PFU holographic videos and 1169 negative holographic videos were collected for training the PFU decision neural network. This dataset was further augmented to create a total of 2594 positive and 3028 negative holographic videos (see the Method sections), where each frame had 480×480 pixels, and the time interval between two consecutive holographic frames was 1 hour.

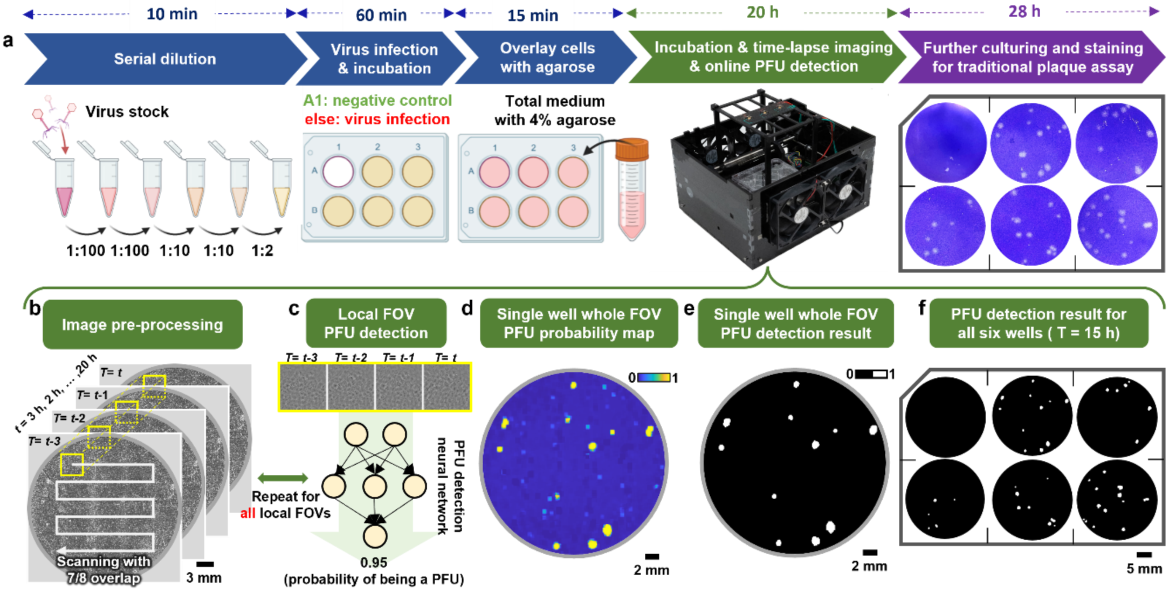


Figure 8.2 Schematics of the workflow of the label-free viral plaque assay and its comparison to the standard PFU assay. (a) Plaque assay sample preparation workflow. The traditional plaque assay at the last step in (a) is only performed for comparison purposes and is not needed for the operation of the presented PFU detection device. (b-f) Detailed image and data processing steps for the live viral plaque assay.

After the neural network-based PFU classifier was trained, it was blindly tested on all the 30 test wells in a scanning manner (shown in Fig. 8.2b) without the need for the PFU localization algorithm, which was only used for the training data generation. For each test well, we have ~18000×18000 effective pixels (representing a 30×30 mm^2 active area after discarding the edges); the digital processing of each test well using the PFU classifier network takes ~7.5 min, which automatically converts the holographic phase images of the well into a PFU probability map (Fig. 8.2d). Each pixel of the well on this map indicates the statistical probability of the local area (0.8×0.8 mm^2) centered at this pixel having a PFU. Using a probability threshold of 0.5 (50%), the final PFU detection and quantification result was obtained across the entire test well area (see e.g., Fig. 8.2 e-f).

Figure 8.3a shows examples of our device's output after 15 hours of incubation. Three representative PFUs are also selected and shown in Fig. 8.3b. When a PFU is in its early stage of growth, with its size much smaller than our $0.8\times0.8\ \mathrm{mm}^2$ virtual scanning window, it appears as a square (shown by the PFU① in Fig. 8.3b) in the final detection result, which effectively is the 2D spatial convolution of the small scale PFU with our scanning window. As another example, PFU③ shows a cluster forming event where the two neighboring PFUs can be easily differentiated using our method as opposed to the traditional plaque assay where they physically merged into one. Fig. 8.3c further shows the PFU quantification achieved by our device compared to the 48-hour traditional plaque assay results. We achieved a detection rate of >90% at 20 hours of incubation *without having any false positives at any time point* despite using no staining.

The presented device is cost-effective, compact, and automated, and can also handle a larger virus concentration range with a more reliable PFU readout. To demonstrate this, we prepared another 5 titer test plates, where for each plate, all the 6 wells were infected by VSV, but with a 2 times dilution difference between each well, covering a large dynamic range in virus concentration from one test well to another. As shown in Fig. 8.4, our method is effective even for the higher virus concentration cases; see, for example, the dilution cases of $2^{-2}\times10^{-4}$ and $2^{-3}\times10^{-4}$. In the traditional 48-hour plaque assay, only the lowest virus concentration is suitable for the PFU quantification due to significant spatial overlapping, whereas for our label-free device, we can automatically and accurately count the individual PFUs at an early stage, even for the highest virus concentration (see Fig. 8.4c).

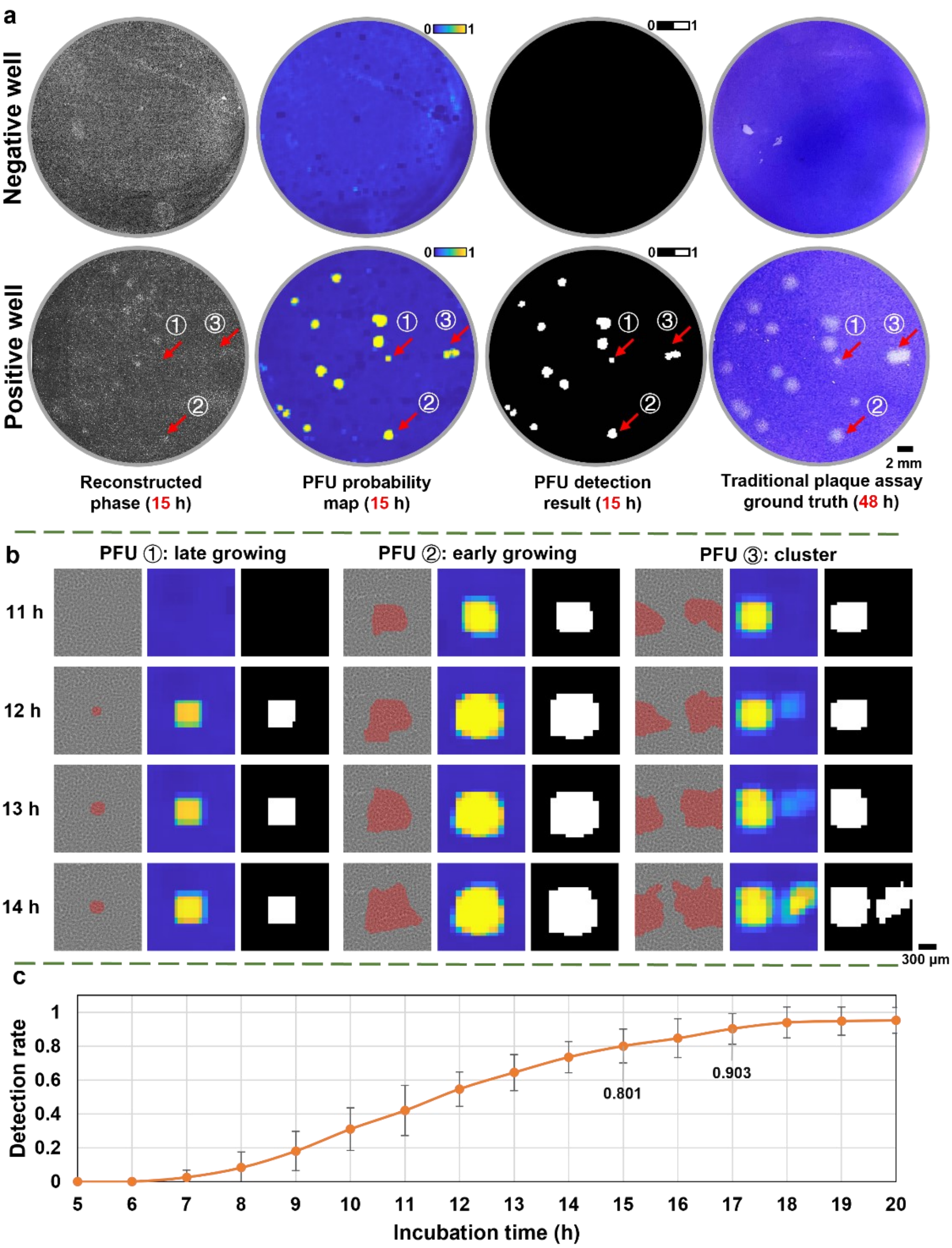


Figure 8.3 Performance of the stain-free plaque assay for samples with low virus concentration. (a) Whole well comparison of the stain-free viral plaque assay after 15-h incubation against the traditional plaque assay after 48-h incubation and staining. (b) The growth of three featured PFUs in the positive well from (a). The reconstructed phase channel is overlaid with the mask generated using the PFU localization algorithm to reveal their locations better. (c) Average PFU detection rate using the label-free viral plaque assay. The error bars show the standard deviation across the 5 testing plates.

Furthermore, our method provides a more reliable readout; for example, in the circled region in Fig. 8.4 a-b, the absence of the cells was caused by some random cell viability problems that occurred during the plaque assay. In our device, these artifacts can be easily differentiated from the cell lysing events caused by the viral replication, since the spatio-temporal patterns for these two events are vastly different (assessed by the trained PFU probability network). This makes our deep learning-enabled device resilient to potential artifacts or cell viability issues randomly introduced during the sample preparation steps.

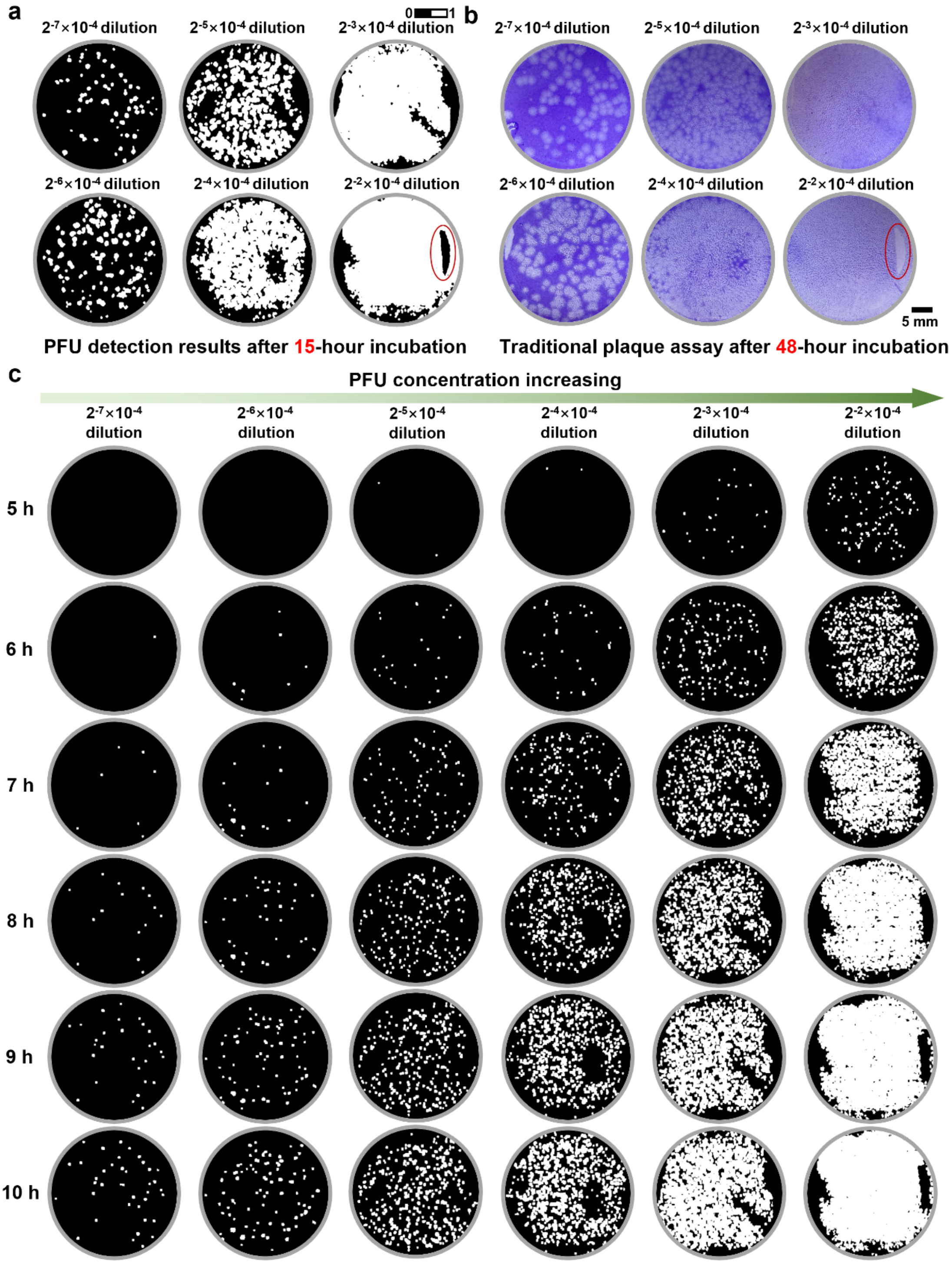


Figure 8.4 Performance of the stain-free viral plaque assay as a function of the virus concentration. (a-b) Whole plate comparison of the stain-free viral plaque assay after 15-h incubation against the traditional plaque assay after 48-h incubation and staining. (c) The growth of PFUs in their early stage for the same plate shown in (a) and (b).

Due to the high virus concentration used in these 5 titer test samples, PFUs quickly clustered and were no longer suitable for manual counting, as shown in Fig. 8.5a. However, the quantitative readout and the PFU probability map of our presented device allowed us to obtain the area of the virus-infected regions across all the time points during the incubation period, as shown in Fig. 8.5b. To better illustrate this, we plot in Fig. 8.5c the virus dilution factor vs. the ratio of the infected cell area per test well (in %) for all the samples at 6, 8, and 10 h of incubation time. Despite the existence of some serial dilution errors, late virus wakeups, and PFU clustering events, the infected area percentage that our device measured is monotonically decreasing with the increasing dilution factor for all the incubation times. This suggests that, by calibrating the system, the virus concentration (PFU/mL) can also be estimated from the percentage of the infected cell area per well.

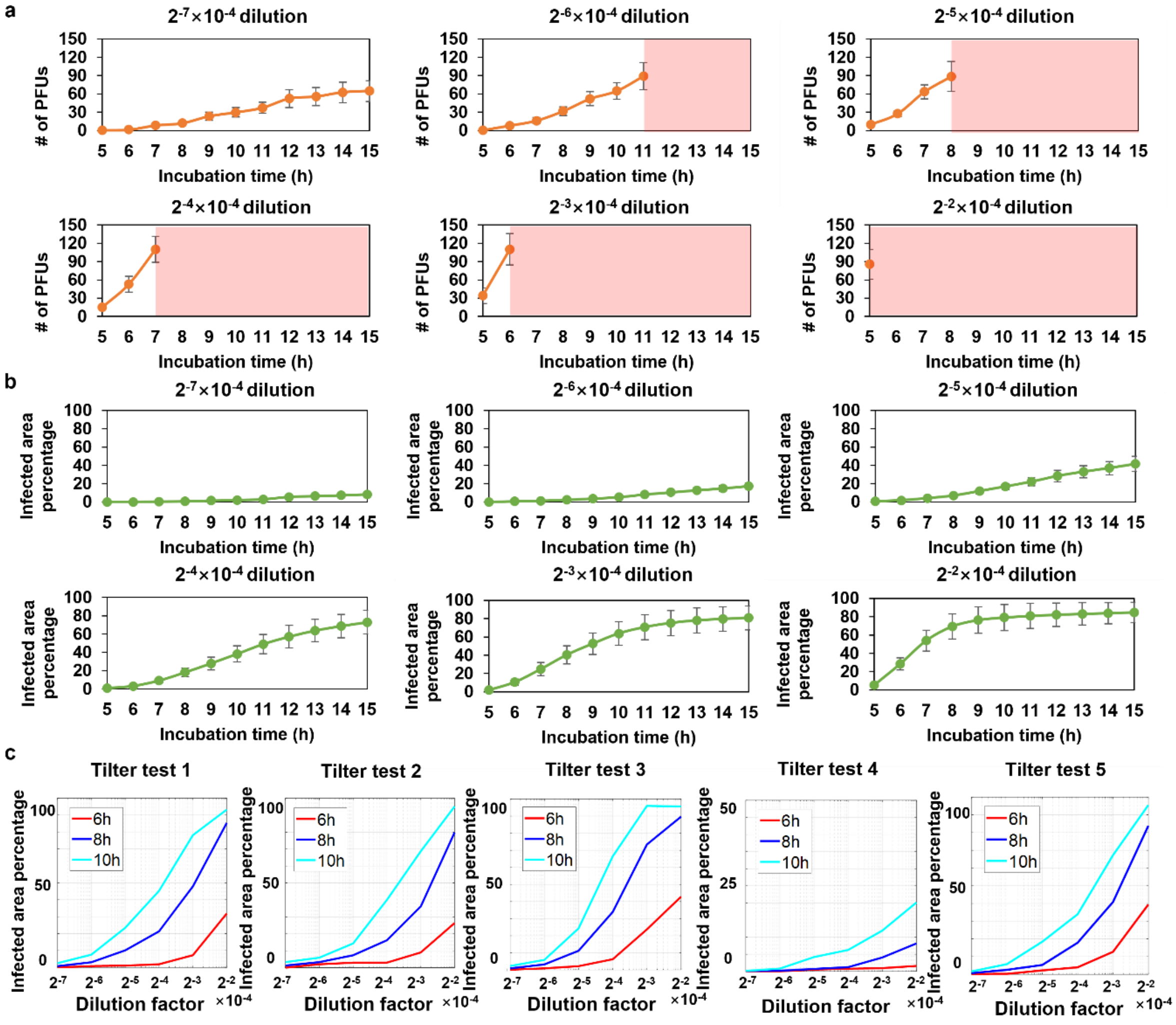


Figure 8.5 Quantitative performance analyses of the label-free viral plaque assay for high virus concentration samples. (a) PFU counting results for different high concentration virus samples at different time points. The light red region indicates the time when the PFUs were heavily clustered and no longer suitable for counting. (b) Area of the virus-infected regions for different high virus concentration samples at different time points. The error bars in (a-b) show the standard error across 5 titer testing plates. (c) Plots of virus dilution factor vs. the ratio of the infected cell area per test well (in %) for all 5 titer test samples at 6, 8, and 10 h of incubation time.

Furthermore, using the area percentage of the virus-infected region as a label-free quantification metric, the presented framework can provide earlier PFU readouts. To show this, we computed the infected area percentage for all the 25 positive/infected wells of the blind testing plates used to generate Fig. 8.3c. As shown in Fig. 8.6, when the infected area percentage is sufficiently large (>1%), a faster PFU concentration readout can be provided at 12-h or 15-h. Since the size of an average PFU on the well is physically larger at 15 hours of incubation compared to 12 hours, the slope of the red calibration curve in Fig. 8.6b is smaller than Fig. 8.6a, as expected. For samples with even higher virus concentrations, the infected cell area percentage could reach >1% in ⩽10 hours of incubation (shown in Fig. 8.5c), providing the PFU concentration readout even earlier.

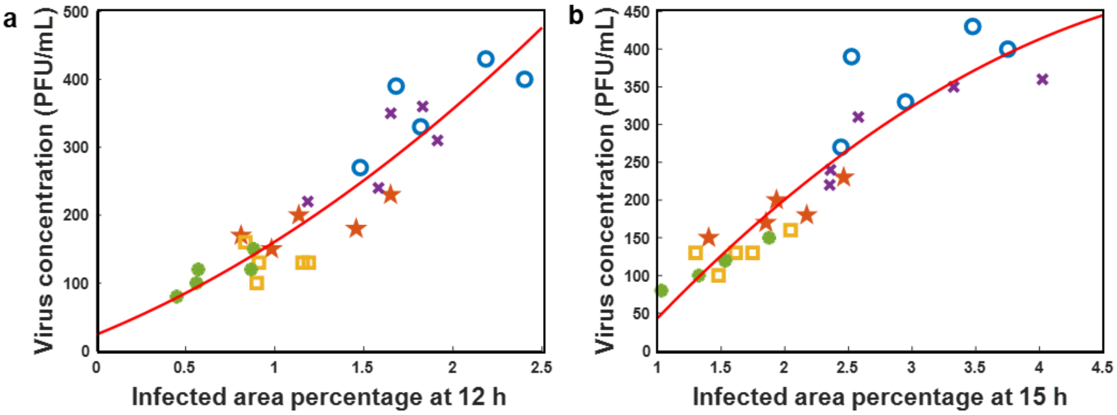


Figure 8.6 Infected area percentage (%) measured by our stain-free device at different time points vs. the virus concentration per well (PFU/mL). The virus concentrations in *y*-axis were

obtained from the 48-h traditional plaque assay for each test well. Different test wells of the same plate were marked with the same color/symbols. There are 25 infected test wells in each plot. The red calibration curves were obtained by quadratic polynomial fitting.

8.3 Discussion

We demonstrated a cost-effective and automated early PFU detection system using a lensfree holographic imaging system and deep learning. This deep learning-based stain-free device captures time-lapse phase images of a test well at a throughput of ~0.32 Giga-pixels per scan, which is then processed by a PFU quantification neural network in ~7.5 min to yield the PFU distribution of each test well. The high detection rate of this label-free device with 100% specificity shown in Fig. 8.3c is a conservative estimate since the ground truth data were obtained after 48-h of incubation. In the early stages of the incubation period, many PFUs did not even exist physically, which means that if we were to use the existing PFUs as the ground truth for our quantification at each time point, our detection rate would be even higher.

The modular design employed by the presented PFU detection platform brings the potential for further system improvements. For example, parallel imaging can be achieved by installing several image sensors on the same system without significantly increasing the cost of the device, which will further improve the 30 cm^2/min effective imaging throughput of the device [301]. More accurate scanning stages can also help reduce the image registration steps needed during image pre-processing. Multi-wavelength phase recovery [302] can also be implemented to improve the overall image quality of the label-free plaques. The presented deep learning-enabled PFU detection framework can be potentially adapted to other imaging modalities that can provide the spatio-temporal differences in the PFU regions for various types of viruses; similarly, the trained PFU classifier network also has the adaptability to these system changes (see the Supplementary Note 8.1).

All in all, we presented a stain-free, rapid, and quantitative viral plaque assay using deep learning and holography. The presented compact and cost-effective device preserves all the advantages of the traditional plaque assays while significantly reducing the required sample incubation time in a label-free manner, saving time and eliminating staining. It is also resilient to potential artifacts during the sample preparation, and can automatically quantify a larger dynamic range of virus concentrations per well. We expect this technique to be widely used in virology research, vaccine development, and related clinical applications.

8.4 Materials and methods

Safety practices: We handled all the cell cultures and viruses during our experiments at our biosafety level 2 (BSL2) laboratory according to the environmental, health, and safety rules and regulations of the University of California, Los Angeles. All operations were carried out under strict aseptic conditions.

Studied organisms: We used Vero C1008 [Vero 76, clone E6, Vero E6] (ATCC® CRL-1586™) (ATCC, USA) and vesicular stomatitis virus (ATCC® VR-1238™). Vero E6 cells are African green monkey kidney cells and are epithelial cells.

Cell propagation: We placed the frozen stock culture immediately in the liquid nitrogen vapor, until ready for use, just after the delivery of the frozen stock culture from ATCC. ATCC formulated Eagle's Minimum Essential Medium (EMEM) (product no. 30-2003, ATCC, USA) was used as a base medium for the cell line. For the complete growth medium, the base medium was mixed with fetal bovine serum (FBS) (product no. 30-2021, ATCC, USA) with a final concentration of 10 %. The FBS stock was aliquoted into 4 mL microcentrifuge tubes and stored at -20°C until use.

We used tissue culture flasks (75 cm^2 area, vented cap, TC treated, T-75) (product no. FB012937, Fisher Scientific, USA) for cell culturing. The base medium in a T-75 flask and FBS were brought to 37°C in the incubator (product no. 51030400, ThermoFisher Scientific, Waltham, MA, USA) and fed with 5% CO_2 before handling it for cell culturing steps. The complete growth medium was prepared. The frozen cell culture was removed from liquid nitrogen and thawed under running water. After thawing the cells, the cell suspension was added to a T-75 flask containing 8 mL of complete growth medium (i.e., EMEM + 10% FBS). The flask was incubated at 37ºC and 5% CO_2 in the incubator. The adherence of the cells to the flask surface was analyzed daily under a phase-contrast microscope. The medium in the flask was renewed 2-3 times a week. The cells were sub-cultivated in a ratio of 1:4 when 95% confluency of the cells as a monolayer was reached.

Subculturing of cells: After the removal of the medium from the cell culture flask, the cells were exposed to 2-3 mL of 0.25% Trypsin/0.53 mM EDTA (ATCC® 30-2101™, ATCC, USA) per flask for dissociation of cell monolayers. The flasks were kept in the incubator for 5-6 minutes for rapid dissociation of cells. 8 mL of complete medium per flask was added to each of them and 2-3 mL of the mixture containing suspended cells was transferred into a new T-75 flask. 8 mL of complete medium was added to the new flask and after gentle mixing, it was incubated at 37ºC and 5% CO_2 for the growth of new cells.

Virus propagation: After the delivery of the virus stock sample, it was stored at 4ºC until use. Virus propagation requires to have Vero cells to be cultured and reach 90-95% confluency on the day of infection. Therefore, Vero cells were cultured 1-2 days before the virus propagation using a seed cell suspension of Vero cells that were subcultured more than 3 times. The virus titer for the stock solution that we had was $10^{6.4}$ $TCID_{(50)}$/mL, which was equal to ~1.75×10^{6} PFU/mL, according to its specifications. The multiplicity of infection was chosen as 0.005, which is recommended to be between 0.005 and 0.00001 [303].

The growth medium in the Vero cell culture flask was removed and discarded. It was rinsed using 5 mL Dulbecco's Phosphate Buffered Saline (D-PBS), 1X (ATCC® 30-2200™) (product no. 30-2200, ATCC, USA). After incubation of the D-PBS containing flask for 3 min in the cabinet, the buffer solution was removed and discarded. 14 µL of stock virus and 6 mL of EMEM (without FBS) were added to each flask. The flasks were incubated at 37°C for 1 hour and rocked at 15 min intervals to have a uniform spread of virus inoculum. After 1 hour, 10 mL of complete medium was added to each flask and the flasks were incubated at 37°C and 5% CO_2 for 48 h to 72 h.

After the incubation, the flasks were analyzed under a phase-contrast microscope. The cells should dissociate from the surface and round cells should be observed in the mixture if the virus propagation process is successful. The mixture was collected into a 50 mL tube (product no. 06-443-20, Fisher Scientific, USA) and the tubes were sealed using a parafilm layer. The suspension in the tube was centrifuged at ~2600 g for 10 min using a centrifuge with swing-out rotors (product no. 22500126, Fisher Scientific, USA). The supernatant containing the virus was collected from the tube and pooled in a new tube. After gentle mixing of the tube to have a uniform suspension, the suspension was aliquoted into 1 mL cryogenic vials with O-ring (product no. 5000-1012, Fisher Scientific, USA). The tubes were labeled and stored in liquid nitrogen.

Preparation of agarose solution: 4% Agarose (product no. MP11AGR0050, Fisher Scientific, USA) in reagent grade water (product no. 23-249-581, Fisher Scientific, USA) was prepared and well mixed [303]. The suspension was then aliquoted into the glass bottles. The solution was sterilized at 121°C for 15 min in an autoclave and 50 mL aliquots were stored at 4°C until use.

Preparation of agarose overlay solution: One of the tubes containing the 50 mL of sterile agarose solution was heated up in a microwave oven for ~30 s. The solution was cooled down to 65°C in a water bath. 23.9 mL EMEM medium was mixed with 0.6 mL FBS and warmed to

50ºC. 3.5 mL of agarose solution were added into the warmed medium mixture using a 10 mL-serological pipette and kept at 50ºC until use.

Well plate preparation: First, the adhered cells in the flask were resuspended using trypsin. The solution was gently mixed to have uniform cell suspension and 10 µL of the suspension was taken for cell counting using a hemacytometer chamber. The cells were counted using a phase-contrast microscope. According to the cell count, the concentration of cells was adjusted to ~6.5×10^5 cells /mL by diluting the suspension using the complete medium. ~6.5×10^5 cells were added to each well of a new 6-well plate. Then, 2 mL of complete medium was added to each well and the plate was stored at 37ºC and 5% CO_2 for 24 h. Next, the cell coverage on each well was checked under the microscope. The cell coverage should reach ~95% to perform the PFU assay. The cells were infected with diluted virus suspension and the overlay solution was added to the cells. After the solidification of the overlay at room temperature, the plate was incubated for 48 h, where the first 20 hours were used for the lensfree PFU imaging device.

Preparation of crystal violet solution: 0.1 g of crystal violet powder (product no. C581-25, Fisher Scientific) was mixed with 40 mL reagent grade water in a 50 mL centrifuge tube. The mixture was gently mixed to dissolve the powder. 10 mL methanol (product no. A452-4, Fisher Scientific) was added to the mixture and stored at room temperature.

Fixation and staining of cells: These steps were only performed for comparison against our device's PFU readings. After 48 h of incubation, the plate was removed from the incubator and the cells were fixed using 0.5 mL methanol/acetic acid solution for 30 min. After 30 min, the wells were washed with water gently to remove the agarose layer. The excess water was removed, and 1 mL of crystal violet (CV) solution was added to each well. The plate with CV was placed into the shaker incubator and mixed at 100 rpm for 30 min. After 30 min of incubation, tap water was used to remove excess stain from the plate and the waste was collected into a large beaker. The plate was left to dry in a fume hood and stored at room temperature by covering with an aluminum foil.

Lens-free imaging set-up: An automatic lens-free PFU imaging set-up was built to capture the in-line holograms of the samples. This set-up includes: 1) a holographic imaging system, 2) a 2D mechanical scanning platform, 3) a cooling system, 4) a controlling circuit, and 5) an automatic controlling program. Three green laser diodes (515 nm, Osram PLT5510) were used for coherent illumination, where each laser diode illuminates two wells on the same column of the 6-well sample plate. The laser diodes were controlled by a driver (TLC5916, Texas Instruments, Texas, US) and mounted ~16 cm away from the sample. A CMOS image sensor (acA3800-14 µm, Basler AG, Ahrensburg, Germany, 1.67 µm pixel size, 6.4 mm × 4.6 mm FOV) was placed ~5 mm beneath the sample forming a lensfree holographic imaging system. The phase changes in the PFU regions were encoded in the acquired holograms.

The FOV of the CMOS image sensor is ~0.3 cm^2, hence mechanical scanning is needed for imaging the whole area of a 6-well plate. A scanning platform was built using a pair of linear translation rails, a pair of linear bearing rods, and linear bearings. 3D printed parts were also used to aid with housing and joints. Two stepper motors (product no. 1124090, Kysan Electronics, San Jose, CA, USA), driven by two driver chips (DRV8834, Pololu Las Vegas, NV, US), were exploited to enable the CMOS sensor to perform 2D horizontal movement. This low-cost platform carries the CMOS sensor moving in a raster pattern and images a total of 420 holograms (21 horizontal, 20 vertical, with 15% overlap) in ~3 min to complete the whole sample scanning. The selected CMOS sensor could heat up to >70°C during its operation, which could disturb the growth of the sample and vaporize the agarose layer, especially for regions that are near the sensor parking location between successive holographic scans. Hence, a cooling system was built using fans (QYN1225BMG-A2, Qirssyn, China). We also sealed the sides of the sample using parafilm (product no. 13-374-16, Fisher Scientific, Hampton, NH, USA) and opened 4 holes on the top cover to form a gentle ventilation system.

A microcontroller (Arduino Micro, Arduino LLC) was used to control the two stepper motor driver chips, the illumination driver chip, and a field-effect transistor-based digital switch (used to turn the CMOS sensor on/off). All these chips along with the digital switch, wires, and capacitors, were integrated on one printed circuit board (PCB), powered by a 6V-1A power adaptor connected to the wall plug.

An automatic controlling program with a graphical user interface (see Supplementary Figure 8.2) was developed using the C++ programming language. It can be used to adjust the image capture parameters (e.g., exposure time etc.) of the CMOS image sensor and communicate with the microcontroller to further switch the laser diodes or CMOS sensor on/off and control the movement of the mechanical scanning system.

All the components along with their unit prices are also summarized in Table S8.1. The cost of the parts of this entire imaging system is < \$880, excluding the laptop computer. At higher volumes of manufacturing, this cost can be significantly reduced.

Image pre-processing: After the image acquisition for each time interval, the raw holograms were firstly reconstructed using the angular spectrum approach based back-propagation [254–258]. The accurate sample-to-sensor distance was estimated at the central region of each well using an auto-focusing algorithm based on the Tamura-of-Gradient (ToG) metric [304] . The same sample-to-sensor distance was used for the entire well since the neural network-based method can tolerate de-focusing. Then, the phase channel of the reconstructed holograms was stitched into the whole FOV image using a correlation-based method and linear blending [225].

Starting from the second time interval, a 2-step image registration was performed to compensate for the low accuracy of the mechanical scanning stage. A coarse whole FOV correlation-based image registration was firstly performed, then a local fine elastic image registration was followed [305].

Coarse PFU localization algorithm

First, each current frame was stacked with the previous 3 frames (shown in Supplementary Figure 8.3a) and a background image (shown in Supplementary Figure 8.3b) was estimated through singular value decomposition [306]. By subtracting this background image, signals from the static regions were suppressed (shown in Supplementary Figure 8.3c). Then, by applying bilateral filtering, the PFU regions with high spatial frequency features were further enhanced (shown in Supplementary Figure 8.3d).

93 images patches (256 × 256 pixels) in PFU regions and 93 image patches from non-PFU regions were cropped manually from 3 experiments. Each pixel of these image patches was labelled as 1 for the PFU region and 0 for the non-PFU region. A Naïve Bayes pixel-wise classifier was trained using this dataset, where the Tamura-of-Gradient (ToG) metric [304] was computed at 2×, 4×, 8×, 16×, and 32× down-sampling scales to serve as the manually selected features. The effect of this classifier is shown in Supplementary Figure 8.3e. Finally, by applying several morphological operations (such as image close, image fill, etc.), the PFU regions are coarsely located (shown in Supplementary Figure 8.3f).

Though this coarse PFU localization algorithm was still subject to detect false positives (shown in Supplementary Figure 8.3g), it could significantly simplify the effort needed for populating the network training dataset. In addition, applying this algorithm to a negative well would help delineate the potential false positives during network training (shown in Supplementary Figure 8.3h). Ultimately, this coarse PFU localization algorithm helped label 357 positive videos and 1169 negative videos used to train the PFU classification network. The positive videos were populated to 2594 by performing augmentation over time; the negative videos were populated to 3028 by further random selection from the negative control wells. Important to note that this

PFU localization algorithm was only used for the training data generation, and was not employed in the blind testing phase as its function was to streamline the training data generation process to be more efficient.

Network architecture and training schedule

Our PFU classifier network was built based on the DenseNet [260] structure, with 2D convolution layers replaced by the pseudo-3D building blocks [261]. The detailed architecture is shown in Supplementary Figure 8.4 and described in Supplementary Note 8.2. We used ReLU as the activation function. Batch normalization and dropout with a rate of 0.5 were used in the training. The loss function we used was the weighted cross-entropy loss:

$$l(p,g)=\sum_{k=1}^{K}\sum_{l=1}^{2}-w_l\cdot g_{k,l}\cdot\log\left(\frac{\exp(p_{k,l})}{\exp(p_{k,1})+\exp(p_{k,2})}\right) \tag{65}$$

where p is the network output, which is the probability of each class (i.e., PFU or non-PFU) before the SoftMax layer, g is the ground-truth label (which is equal to 0 or 1 for binary classification), K is the total number of training samples in one batch, w is the weight assigned to each class, defined as w = 1-d, where d is the percentage of the samples in one class (d = 46.1% for positive class, d = 53.9% for negative class).

The input 4-frame videos were formatted as a tensor with the dimension of 1 × 4 × 480 × 480 (channel × time frame × height × width). Data augmentation, such as flipping, and rotation were applied when loading the training dataset. The network model was optimized using the Adam optimizer with a momentum coefficient of (0.9, 0.999). The learning rate started as 1×10^{-4} and a scheduler was used to decrease the learning rate with a coefficient of 0.7 at every 30 epochs. Our model was trained for 264 epochs using NVIDIA GeForce RTX3090 GPU with a batch size of 30. The loss curve, training sensitivity and specificity curves of our training process are provided in Supplementary Figure 8.5. In these curves, 10% of the training dataset was randomly selected as the validation dataset. Note that the training and validation datasets (containing holographic videos of the wells) were formed from various wells at different time points of each PFU assay as detailed earlier; therefore, these training and validation sensitivity and specificity curves do not reflect the evaluation of an individual test well that is periodically monitored from the beginning of the incubation. Our blind testing results reported in the Results section, however, were acquired by using the trained PFU detection neural network on individual test wells that were continuously monitored from the beginning of the incubation with a sampling period of 1 hour, achieving >90% detection rate for PFUs with 100% specificity in <20 hours.

Image post-processing

After getting the PFU probability map and applying the 0.5 threshold, two image post-processing steps were followed to obtain the final PFU detection result: 1) maximum probability projection along time, and 2) PFU size thresholding. The maximum projection was used to compensate for the lower PFU probability values generated from the PFU center when it enters the late stage of its growth. The effect of this maximum projection is illustrated in Supplementary Figure 6. The size threshold on the PFU probability map was set to 0.5×0.5 mm^2.

8.5 Appendix

Supplementary Note 1: Guidelines for hyperparameter selection to adapt to other modalities and biological agents.

The presented framework has the potential to be adapted to different imaging modalities that can provide spatio-temporal differences in the PFU regions for various types of biological agents. Here, we discuss the principles of the system hyperparameter selection, particularly the 0.8×0.8 mm^2 network input window size, and the network input frames, to provide a guideline for future applications.

The selection of the window size should take into account the system resolution, PFU size, and network structure. The input window size to the PFU detection network should be approximately one order of magnitude larger compared to the resolution of the imaging system to provide sufficient spatial information to the network. On the other hand, if the window size is too large, it will dramatically decrease the network inference speed and harm its ability to differentiate PFU clusters at an earlier time point. Lastly, the number of pixels for the window must be divisible by 32, since the selected network structure will down-sample the input images by 32 times; of course, the network structure can be modified accordingly to handle a different number of pixels at the input depending on the needs. Combining all these, the 480×480 pixels, i.e., 0.8×0.8 mm^2 window size was chosen in our PFU detection network.

For the selection of the number of input frames, our experience is that at least 3 time-lapse frames must be fed into the network to differentiate an early-stage PFU from other non-specific signals. To increase the stability of the network performance and ensure its efficiency, we used 4 frames (acquired at a period of 1 hour) as our network input. However, this number is subject to increase when the 1-h scanning time interval is reduced. This should be ultimately decided by whether sufficient spatio-temporal features can be captured when adapting to different types of viruses depending on the corresponding plaque formation speed.

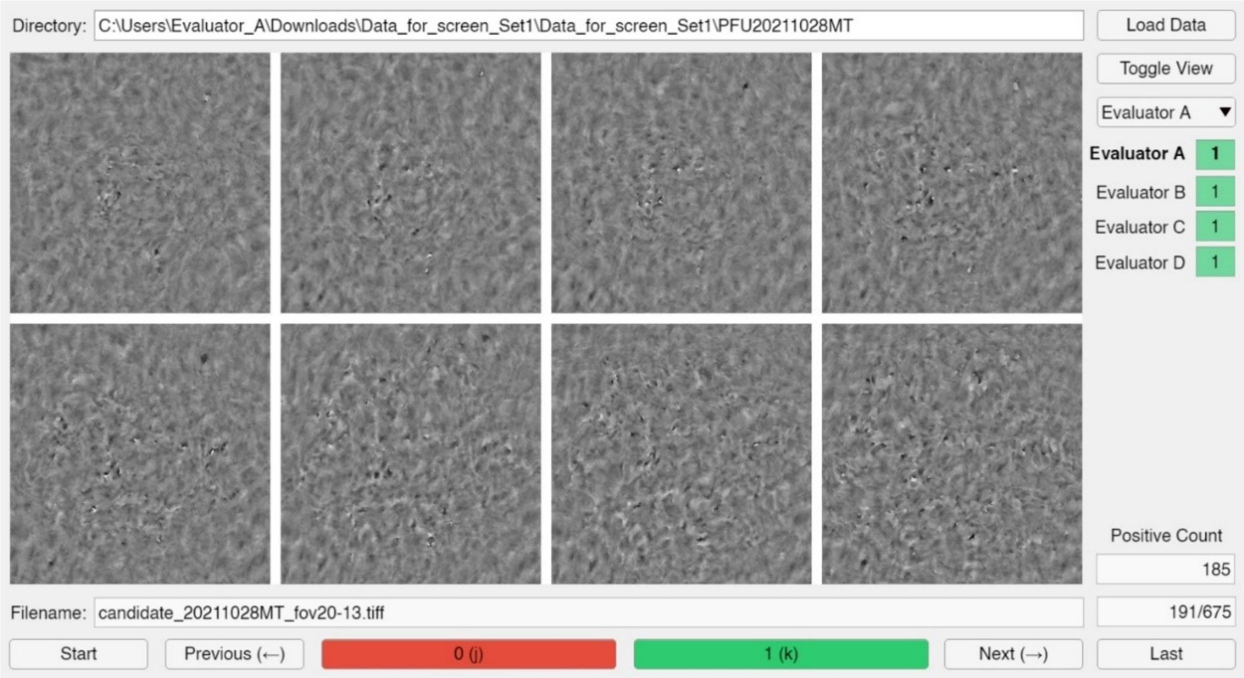
Supplementary Note 2: *Network architecture of the PFU decision neural network.*

Dense layers and transition layers were connected alternatively to transfer the high-dimensional data into low-dimensional data. For each dense layer taking 4-dimensional input, a 3D convolutional layer with a kernel size of (1,1,1) and a stride of (1,1,1) was first applied to reduce the number of channels to 32, and the pseudo-3D block (sequentially selected from the 3 types shown in Supplementary Figure 4) was followed to further extract both the spatial and temporal features, the output of which was then concatenated with the original input, following the structure of DenseNet. Each transition layer also reduced the number of channels of its input by half using a 3D convolutional layer with a kernel size of (1,1,1) and a stride of (1,1,1), and then an average pooling layer with a kernel size of (2,2,2) and a stride of (2,2,2) was followed to reduce the image size by half. We also included 2D versions of our dense and transition layers, which were only used to process the spatial domain in the case when the temporal dimension was collapsed to 1. In total, our network consists of 6 3D dense layers, 2 3D transition layers, 15 2D dense layers, and 1 2D transition layer. Lastly, we had an average spatial pooling layer with a kernel size of (15,15) to flatten the features into a 113-length vector, which was then fed into a fully connected layer and SoftMax layer to produce an output PFU probability map.

Table S8.1 Cost of our label-free holographic PFU imaging device with the unit price of each component under low volume.

Component	Unit price	Number	Total price
515 nm laser diode	\$15.36	3	\$46.08
Basler CMOS sensor	\$584.00	1	\$584.00
Stepper motor	\$14.00	2	\$28.00
Arduino Micro	\$22.80	1	\$22.80
TLC5916 LED driver	\$1.00	1	\$1.00
Stepper motor driver	\$9.95	2	\$19.9
PCB fabrication	\$2.00	1	\$2.00
Belts	\$25.00	/	\$25.00
Acrylic sheet	\$25.00	2	\$50.00
Mechanical supporting parts	\$50.00	/	\$50.00
3D printing material	\$50.00	/	\$50.00
Total			**\$878.78**

(a)



Directory: C:\Users\Evaluator_A\Downloads\Data_for_screen_Set1\Data_for_screen_Set1\PFU20211028MT

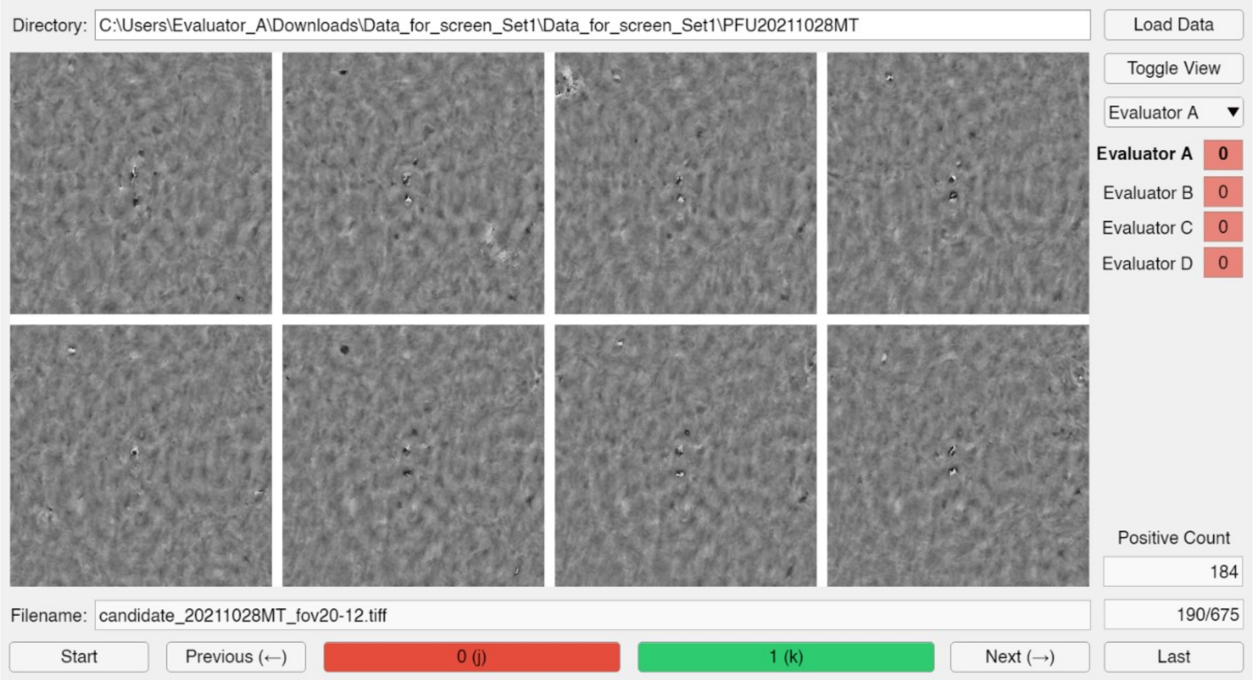
Evaluator A: 1
Evaluator B: 1
Evaluator C: 1
Evaluator D: 1

Positive Count: 185

Filename: candidate_20211028MT_fov20-13.tiff

191/675

(b)



Directory: C:\Users\Evaluator_A\Downloads\Data_for_screen_Set1\Data_for_screen_Set1\PFU20211028MT

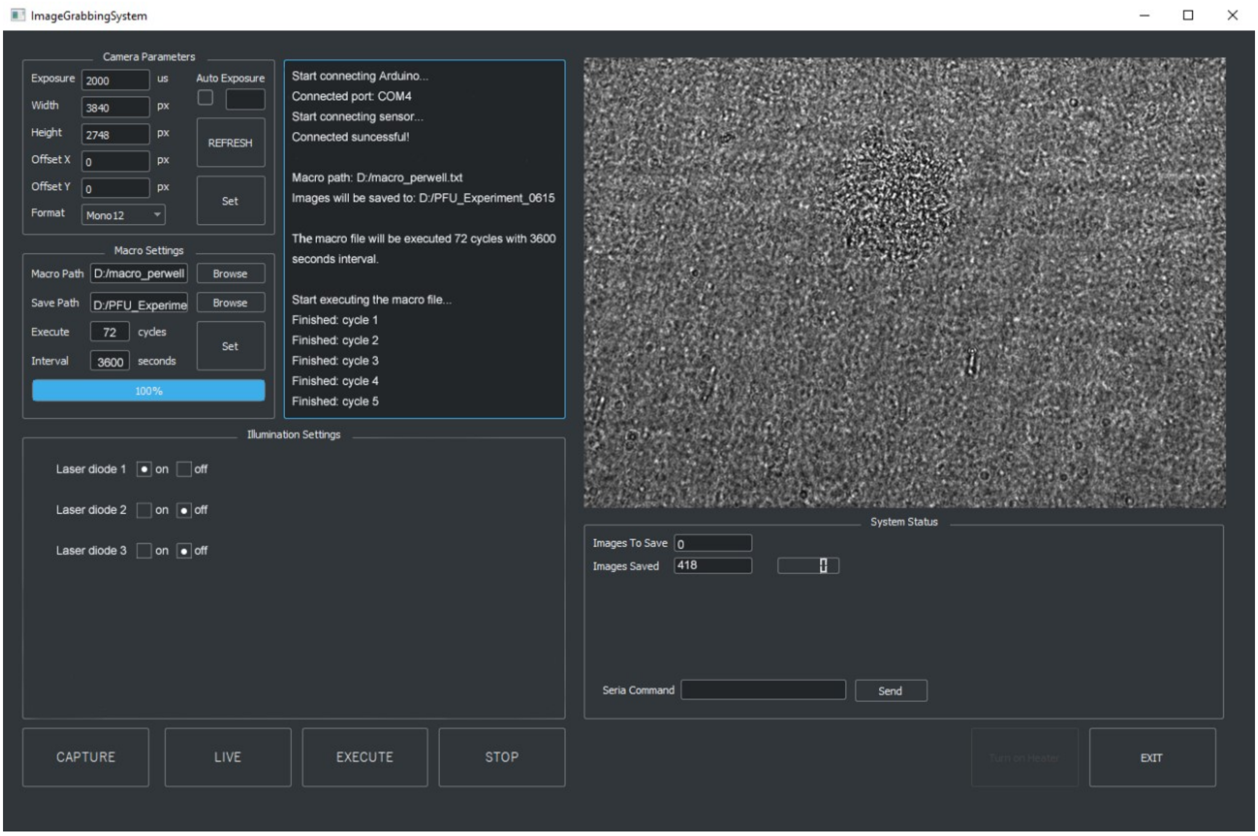
Evaluator A: 0
Evaluator B: 0
Evaluator C: 0
Evaluator D: 0

Positive Count: 184

Filename: candidate_20211028MT_fov20-12.tiff

190/675

Figure S8.1 The Graphical User Interface (GUI) for manually labeling the PFU videos; this was only used during the training phase to efficiently curate labeled image data. (a) Example of a true positive PFU at 1, 3, 5, …, 15 hours. (b) Example of a false positive PFU at 1, 3, 5, …, 15 hours (usually these are bubbles, dust, agar degradation regions, etc.). Undecided PFU candidates were further compared with the 48-h traditional plaque assay (after staining) for confirmation.



Exposure 2000 us
Width 3840 px
Height 2748 px
Offset X 0 px
Offset Y 0 px
Format Mono12

Macro Path D:/macro_perwell
Save Path D:/PFU_Experime
Execute 72 cycles
Interval 3600 seconds
100%

```
Start connecting Arduino...
Connected port: COM4
Start connecting sensor...
Connected suncessful!

Macro path: D:/macro_perwell.txt
Images will be saved to: D:/PFU_Experiment_0615

The macro file will be executed 72 cycles with 3600
seconds interval.

Start executing the macro file...
Finished: cycle 1
Finished: cycle 2
Finished: cycle 3
Finished: cycle 4
Finished: cycle 5
```

Laser diode 1 ● on ☐ off
Laser diode 2 ☐ on ● off
Laser diode 3 ☐ on ● off

Images To Save 0
Images Saved 418
Seria Command

Figure S8.2 Graphical user interface of the imaging system controlling program. Users can adjust the illumination, image sensor, and scan settings through this user interface.

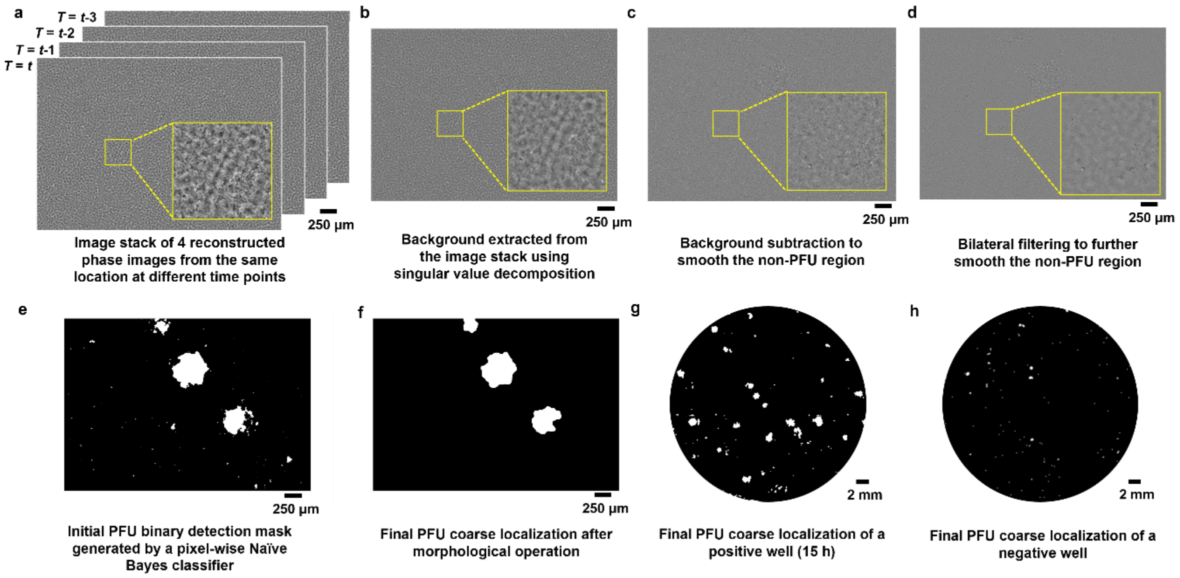


Image stack of 4 reconstructed phase images from the same location at different time points

Background extracted from the image stack using singular value decomposition

Background subtraction to smooth the non-PFU region

Bilateral filtering to further smooth the non-PFU region

Initial PFU binary detection mask generated by a pixel-wise Naïve Bayes classifier

Final PFU coarse localization after morphological operation

Final PFU coarse localization of a positive well (15 h)

Final PFU coarse localization of a negative well

Figure S8.3 Workflow of the coarse PFU localization algorithm, which is only used during the training phase for efficient data curation. A coarse PFU localization mask (binary) can be obtained using the PFU localization algorithm following the steps from (a) to (f).

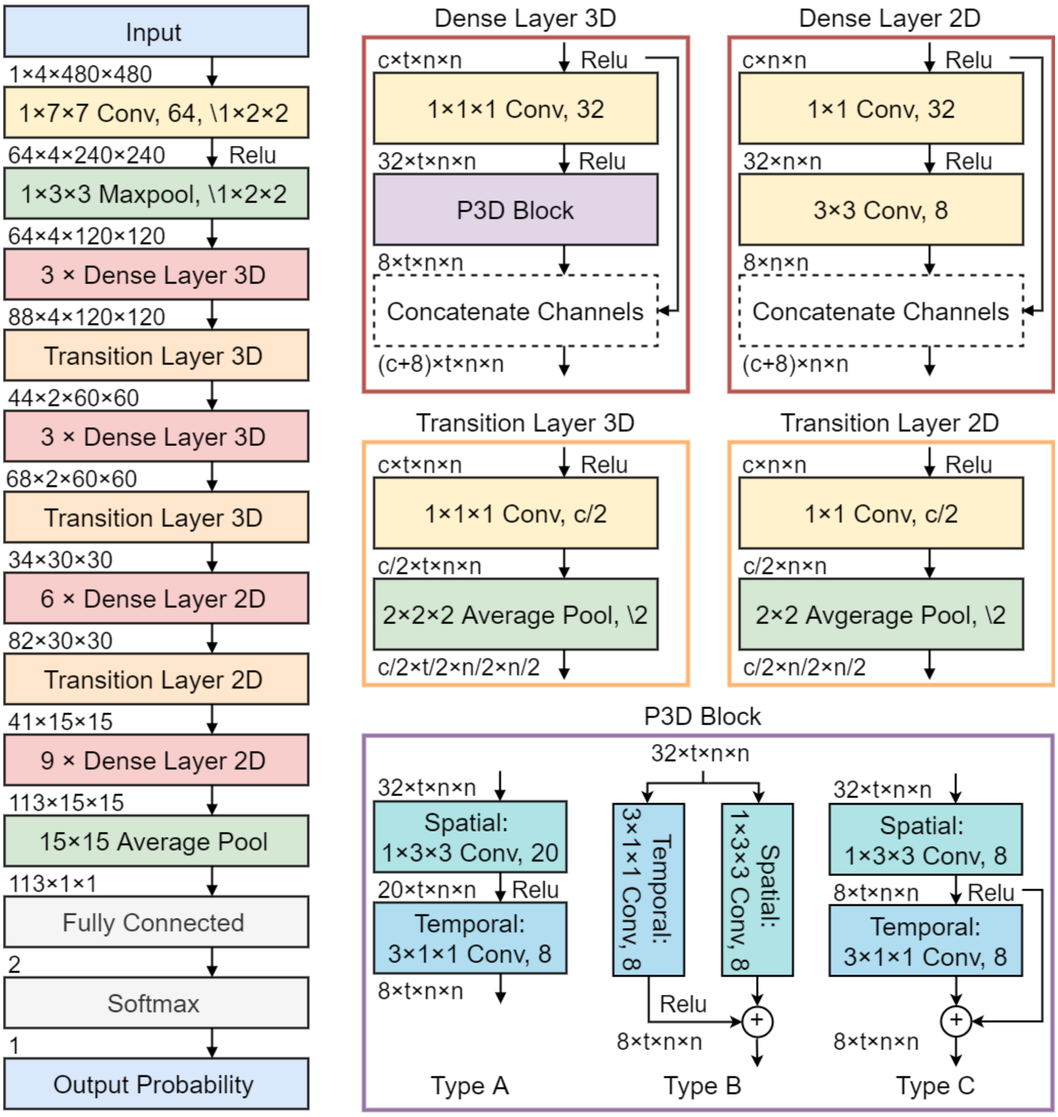


Figure S8.4 Network architecture of the PFU decision neural network. This network is based on the DenseNet structure, with the 2D convolutional layers replaced by the pseudo-3D building blocks.

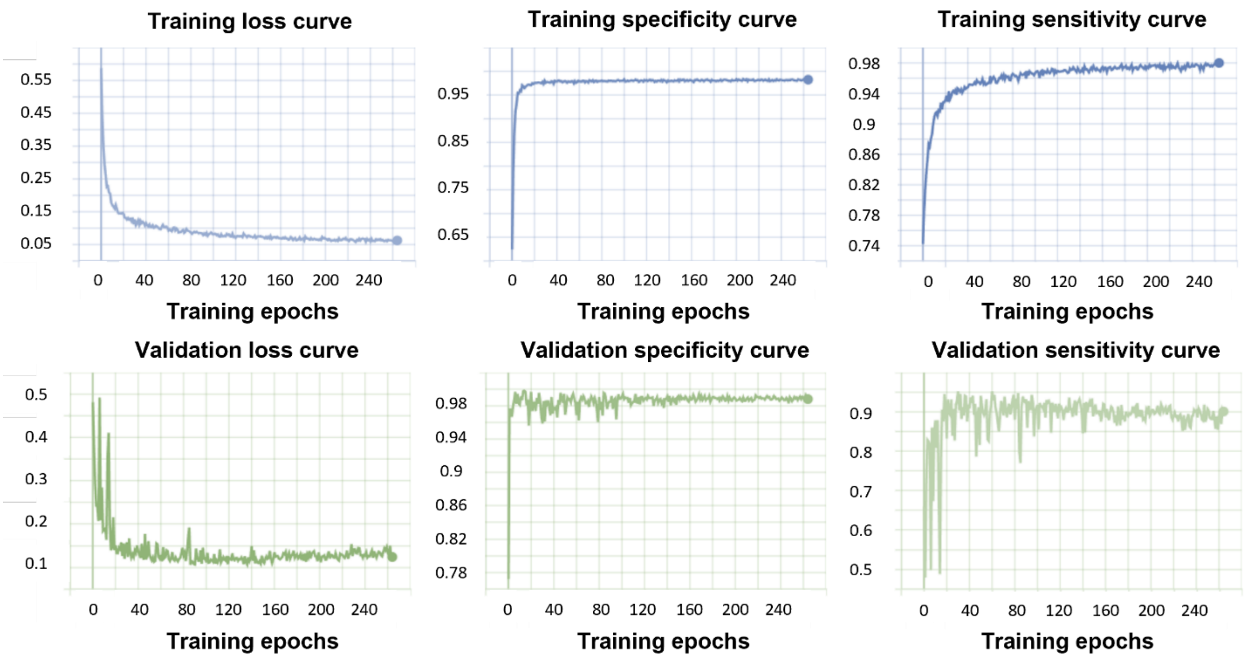


Figure S8.5 Loss, sensitivity, and specificity curves of the PFU decision neural network training process.

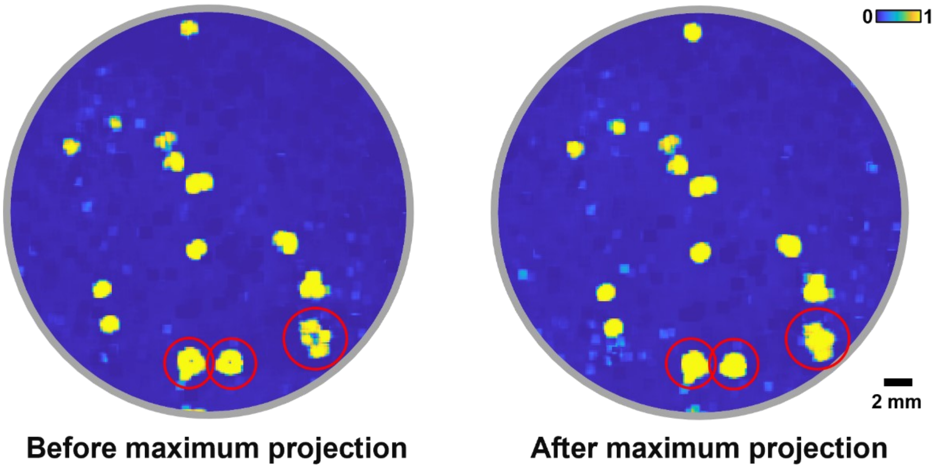


Figure S8.6 Effect of the maximum projection method. The impact of using the maximum projection method to avoid lower PFU probability values being generated from the center of the late-stage PFUs (see e.g., the central regions of the red circled areas). Also see the Methods section of the main text.

Chapter 9 . Conclusions

This dissertation has focused on the use of deep learning in the optical microscopy imaging. Due to the flexibility of deep learning, this technique can not only improve the performance of an existing optical system, but also adapts the optical system to more challenging biomedical imaging applications. To conclude, I summarize my major contributions to deep learning on microscopy imaging in the following domains:

(1) Optics. I developed a deep learning-based technique to improve the resolution for a coherent imaging system and showed its efficacy for both pixel-size-limited case and diffraction limited case, which suggests a great potential for adapting this framework to a wide range of imaging platforms (Chapter 2). Next, I presented the use of deep learning on simultaneously eliminating the missing-phase-related artifact and correct the color inaccuracy due to the use of only three illuminating wavelengths in holography, which reduced the needed data for reconstruct a high-fidelity color image using the lensfree imaging system by >200 folds (Chapter 3). Later, I created a deep learning-enabled holographic polarization microscope, which can reconstruct both the quantitative retardance and orientation information of a birefringent object using a single polarization state in the light path. This framework eliminates the need for altering the light path during imaging and provides a practical solution for polarization imaging using holography (Chapter 4).

(2) Pathology. The development of the deep learning color holography framework can dramatically improve the system throughput for adopting the lenfree imaging technique in pathology (Chapter 3). In addition, I developed a deep learning-based label-free virtual staining technique, which could further accelerate the pathological diagnosing workflow by eliminating the chemical staining process (Chapter 5).

(3) Microbiology. I led the development of a computational cytometer based on magnetically

modulated lensless speckle imaging. Using deep learning, this framework can differentiate the target rare cell to other random particles in the blood sample (Chapter 6). Furthermore, I developed a CFU detection and classification framework using the TFT sensor, which significantly reduces the required sample incubation time with an ultra large imaging FOV and a cost-effective system design (Chapter 7). Last but not the least, I created a novel quantitative viral plaque assay technique that requires no staining and can provide early PFUs readout during their growth (Chapter 8).

Looking forward, I believe that many of these techniques and platforms have the great potential to be translated to rapid, compact, and cost-effective diagnosis tools. Yet, there is still a long way to go for applying deep learning in the field of biomedical imaging. For example, one of the greatest burdens for applying deep learning in such applications is to create an unbiased training dataset. A systematical standardized evaluation method is yet needed in many bio-imaging applications. Further, the interpretation of the black box nature of the deep learning remains to be a high interest for many researchers. Combined with the rapid development of deep learning, there is still a wide range of opportunities to further penetrate this technique into microscopy imaging.

References

1. G. Popescu, *Quantitative Phase Imaging of Cells and Tissues* (McGraw Hill Professional, 2011).
2. D. Gabor, "A New Microscopic Principle," Nature **161**, 777–778 (1948).
3. E. N. Leith and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory*," J. Opt. Soc. Am., JOSA **52**, 1123–1130 (1962).
4. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
5. A. Greenbaum, W. Luo, T.-W. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, and A. Ozcan, "Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy," Nat. Methods **9**, 889–895 (2012).
6. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," Journal of the Optical Society of America **73**, 1434 (1983).
7. L. J. Allen and M. P. Oxley, "Phase retrieval from series of images obtained by defocus variation," Optics Communications **199**, 65–75 (2001).
8. I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," Opt. Lett., OL **22**, 1268–1270 (1997).
9. Z. Wang, L. Millet, M. Mir, H. Ding, S. Unarunotai, J. Rogers, M. U. Gillette, and G. Popescu, "Spatial light interference microscopy (SLIM)," Opt Express **19**, 1016–1026 (2011).
10. W. Luo, Y. Zhang, Z. Göröcs, A. Feizi, and A. Ozcan, "Propagation phasor approach for holographic image reconstruction," Scientific Reports **6**, 22738 (2016).
11. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," (2017).
12. Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Science & Applications **7**, e17141 (n.d.).
13. Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. De Haan, and A. Ozcan, "Cross-modality deep learning brings bright-field microscopy contrast to holography," arXiv:1811.07103 [physics] (2018).
14. Y. Wu, Y. Rivenson, Y. Zhang, Z. Wei, H. Günaydin, X. Lin, and A. Ozcan, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," Optica,

OPTICA **5**, 704–710 (2018).

15. V. Mico, Z. Zalevsky, P. García-Martínez, and J. García, "Superresolved imaging in digital holography by superposition of tilted wavefronts," Appl. Opt., AO **45**, 822–828 (2006).
16. V. Mico, Z. Zalevsky, P. García-Martínez, and J. García, "Synthetic aperture superresolution with multiple off-axis holograms," J. Opt. Soc. Am. A, JOSAA **23**, 3162–3170 (2006).
17. W. Bishara, T.-W. Su, A. F. Coskun, and A. Ozcan, "Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution," Optics Express **18**, 11181 (2010).
18. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nature Photonics **7**, 739–745 (2013).
19. W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," Light Sci Appl. **5**, e16060 (2016).
20. A. Meiri, E. Gur, J. Garcia, V. Micó, B. Javidi, and Z. Zalevsky, "Super Resolved Holographic Configurations," Multi-Dimensional Imaging 225–239 (2014).
21. M. Paturzo, F. Merola, S. Grilli, S. De Nicola, A. Finizio, and P. Ferraro, "Super-resolution in digital holography by a two-dimensional dynamic phase grating," Optics express **16**, 17107–17118 (2008).
22. A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, "Sparsity-based single-shot subwavelength coherent diffractive imaging," Nat Mater **11**, 455–459 (2012).
23. Y. Liu, L. Tian, C.-H. Hsieh, and G. Barbastathis, "Compressive holographic two-dimensional localization with $1/30^2$ subpixel accuracy," Opt. Express, OE **22**, 9774–9782 (2014).
24. Y. Rivenson, M. A. Shalev, and Z. Zalevsky, "Compressive Fresnel holography approach for high-resolution viewpoint inference," Optics Letters **40**, 5606 (2015).
25. C. Fournier, F. Jolivet, L. Denis, N. Verrier, E. Thiebaut, C. Allier, and T. Fournel, "Pixel super-resolution in digital holography by regularized reconstruction," Applied optics **56**, 69–77 (2017).
26. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica, OPTICA **4**, 1437–1443 (2017).
27. Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Gunaydin, Y. Zhang, Z. Gorocs, K. Liang, D. Tseng, and A. Ozcan, "Deep learning enhanced mobile-phone microscopy," ACS Photonics

(2018).

28. W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," Nature Biotechnology (2018).

29. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," Optica, OPTICA **5**, 458–464 (2018).

30. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nature Methods **16**, 103–110 (2019).

31. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2014), pp. 2672–2680.

32. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing **13**, 600–612 (2004).

33. A. Greenbaum, Y. Zhang, A. Feizi, P.-L. Chung, W. Luo, S. R. Kandukuri, and A. Ozcan, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," Science Translational Medicine **6**, 267ra175-267ra175 (2014).

34. A. Greenbaum and A. Ozcan, "Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy," Opt. Express, OE **20**, 3129–3143 (2012).

35. A. Greenbaum, U. Sikora, and A. Ozcan, "Field-portable wide-field microscopy of dense samples using multi-height pixel super-resolution based lensfree imaging," Lab Chip **12**, 1242–1245 (2012).

36. Y. Rivenson, Y. Wu, H. Wang, Y. Zhang, A. Feizi, and A. Ozcan, "Sparsity-based multi-height phase recovery in holographic microscopy," Scientific Reports **6**, 37862 (2016).

37. J. W. Goodman, *Statistical Optics* (John Wiley & Sons, 2015).

38. Y. Zhang, H. Wang, Y. Wu, M. Tamamitsu, and A. Ozcan, "Edge sparsity criterion for robust holographic autofocusing," Opt Lett **42**, 3824–3827 (2017).

39. N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," IEEE Transactions on Pattern Analysis and Machine Intelligence **22**, 831–843 (2000).

40. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv:1505.04597 [cs] (2015).

41. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs] (2014).

42. W. C. Revie, M. Shires, P. Jackson, D. Brettle, R. Cochrane, and D. Treanor, "Color management in digital pathology," Analytical Cellular Pathology **2014**, (2014).

43. W. S. Campbell, G. A. Talmon, K. W. Foster, S. M. Lele, J. A. Kozel, and W. W. West, "Sixty-five thousand shades of gray: importance of color in surgical pathology diagnoses," Human pathology **46**, 1945–1950 (2015).

44. P. A. Bautista, N. Hashimoto, and Y. Yagi, "Color standardization in whole slide imaging using a color calibration slide," Journal of pathology informatics **5**, (2014).

45. P. Shrestha and B. Hulsken, "Color accuracy and reproducibility in whole slide imaging scanners," Journal of Medical Imaging **1**, 027501 (2014).

46. M. S. Peercy and L. Hesselink, "Wavelength selection for true-color holography," Applied optics **33**, 6811–6817 (1994).

47. D. Pascale, "A review of rgb color spaces... from xyy to r’g’b’," Babel Color **18**, 136–152 (2003).

48. Y. Zhang, T. Liu, Y. Huang, D. Teng, Y. Bian, Y. Wu, Y. Rivenson, A. Feizi, and A. Ozcan, "Accurate color imaging of pathology slides using holography and absorbance spectrum estimation of histochemical stains," Journal of Biophotonics e201800335 (2018).

49. S. G. Kalenkov, G. S. Kalenkov, and A. E. Shtanko, "Hyperspectral digital holography of microobjects," in *Practical Holography XXIX: Materials and Applications* (International Society for Optics and Photonics, 2015), Vol. 9386, p. 938604.

50. P. Xia, Y. Ito, Y. Shimozato, T. Tahara, T. Kakue, Y. Awatsuji, K. Nishio, S. Ura, T. Kubota, and O. Matoba, "Digital Holography Using Spectral Estimation Technique," J. Display Technol., JDT **10**, 235–242 (2014).

51. W. Luo, A. Greenbaum, Y. Zhang, and A. Ozcan, "Synthetic aperture-based on-chip microscopy," Light: Science & Applications **4**, e261 (2015).

52. W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," Light Sci Appl. **5**, e16060 (2016).

53. W. Bishara, T.-W. Su, A. F. Coskun, and A. Ozcan, "Lensfree on-chip microscopy over a wide field-

of-view using pixel super-resolution," Optics Express **18**, 11181 (2010).

54. W. Bishara, U. Sikora, O. Mudanyali, T.-W. Su, O. Yaglidere, S. Luckhart, and A. Ozcan, "Holographic pixel super-resolution in portable lensless on-chip microscopy using a fiber-optic array," Lab on a Chip **11**, 1276 (2011).

55. A. Greenbaum and A. Ozcan, "Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy," Optics Express **20**, 3129 (2012).

56. A. Greenbaum, U. Sikora, and A. Ozcan, "Field-portable wide-field microscopy of dense samples using multi-height pixel super-resolution based lensfree imaging," Lab on a Chip **12**, 1242 (2012).

57. S. O. Isikman, W. Bishara, and A. Ozcan, "Lensfree On-chip Tomographic Microscopy Employing Multi-angle Illumination and Pixel Super-resolution," Journal of Visualized Experiments (2012).

58. A. Greenbaum, N. Akbari, A. Feizi, W. Luo, and A. Ozcan, "Field-Portable Pixel Super-Resolution Colour Microscope," PLoS ONE **8**, e76475 (2013).

59. A. Greenbaum, A. Feizi, N. Akbari, and A. Ozcan, "Wide-field computational color imaging using pixel super-resolved on-chip microscopy," Optics Express **21**, 12469 (2013).

60. A. Greenbaum, W. Luo, B. Khademhosseinieh, T.-W. Su, A. F. Coskun, and A. Ozcan, "Increased space-bandwidth product in pixel super-resolved lensfree on-chip microscopy," Sci. Rep. **3**, (2013).

61. Y. Wu, Y. Zhang, W. Luo, and A. Ozcan, "Demosaiced pixel super-resolution for multiplexed holographic color imaging," Sci Rep **6**, (2016).

62. Y. Cao, Z. Zhou, W. Zhang, and Y. Yu, "Unsupervised Diverse Colorization via Generative Adversarial Networks," in *Machine Learning and Knowledge Discovery in Databases*, M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroski, eds., Lecture Notes in Computer Science (Springer International Publishing, 2017), pp. 151–166.

63. A. Deshpande, J. Lu, M.-C. Yeh, M. Jin Chong, and D. Forsyth, "Learning diverse image colorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 6837–6845.

64. Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. de Haan, and A. Ozcan, "Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," Light: Science & Applications **8**, 25 (2019).

65. 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 **8**, 23 (2019).

66. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, and A. E. Sisk, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering 1 (2019).

67. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing **13**, 600–612 (2004).

68. B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula," ACM Transactions on Graphics (TOG) **16**, 109–154 (1997).

69. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).

70. Y. Zhang, H. Wang, Y. Wu, M. Tamamitsu, and A. Ozcan, "Edge sparsity criterion for robust holographic autofocusing," Optics Letters **42**, 3824 (2017).

71. M. Tamamitsu, Y. Zhang, H. Wang, Y. Wu, and A. Ozcan, "Comparison of Gini index and Tamura coefficient for holographic autofocusing based on the edge sparsity of the complex optical wavefront," arXiv:1708.08055 [physics.optics] (2017).

72. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (2014), pp. 2672–2680.

73. T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," Scientific reports **9**, 3926 (2019).

74. K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, "Resolution enhancement in scanning electron microscopy using deep learning," arXiv:1901.11094 [physics] (2019).

75. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv:1505.04597 [cs] (2015).

76. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Science & Applications **7**, 17141 (2018).

77. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks.

2010," in *Internaional Conference on Artificial Intelligence and Statistics* (n.d.).

78. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

79. S. Preibisch, S. Saalfeld, and P. Tomancak, "Globally optimal stitching of tiled 3D microscopic image acquisitions," Bioinformatics **25**, 1463–1465 (2009).

80. M. Wolman, "Polarized light microscopy as a tool of diagnostic pathology.," Journal of Histochemistry & Cytochemistry **23**, 21–50 (1975).

81. P. Arun Gopinathan, G. Kokila, M. Jyothi, C. Ananjan, L. Pradeep, and S. Humaira Nazir, "Study of collagen birefringence in different grades of oral squamous cell carcinoma using picrosirius red and polarized light microscopy," Scientifica **2015**, (2015).

82. B. Vijaya, B. S. Dalal, and G. V. Manjunath, "Primary cutaneous amyloidosis: a clinico-pathological study with emphasis on polarized microscopy," Indian Journal of Pathology and Microbiology **55**, 170 (2012).

83. L.-W. Jin, K. A. Claborn, M. Kurimoto, M. A. Geday, I. Maezawa, F. Sohraby, M. Estrada, W. Kaminksy, and B. Kahr, "Imaging linear birefringence and dichroism in cerebral amyloid pathologies," Proceedings of the National Academy of Sciences **100**, 15294–15298 (2003).

84. G. G. Cornwell III, W. L. Murdoch, R. A. Kyle, P. Westermark, and P. Pitkänen, "Frequency and distribution of senile cardiovascular amyloid: a clinicopathologic correlation," The American journal of medicine **75**, 618–623 (1983).

85. D. J. Mccarty and J. L. Hollander, "Identification of urate crystals in gouty synovial fluid," Ann. Intern. Med. **54**, 452–460 (1961).

86. N. W. McGill and P. A. Dieppe, "Evidence for a promoter of urate crystal formation in gouty synovial fluid," Ann. Rheum. Dis. **50**, 558–561 (1991).

87. R. A. Gatter, "Editorial: The compensated polarized light microscope in clinical rheumatology," Arthritis Rheum. **17**, 253–255 (1974).

88. C. Gordon, A. Swan, and P. Dieppe, "Detection of crystals in synovial fluids by light microscopy: sensitivity and reliability," Ann. Rheum. Dis. **48**, 737–742 (1989).

89. J. W. Park, D. J. Ko, J. J. Yoo, S. H. Chang, H. J. Cho, E. H. Kang, J. K. Park, Y. W. Song, and Y. J. Lee, "Clinical factors and treatment outcomes associated with failure in the detection of urate crystal

in patients with acute gouty arthritis," Korean J. Intern. Med. **29**, 361–369 (2014).

90. R. Oldenbourg and G. Mei, "New polarized light microscope with precision universal compensator," Journal of microscopy **180**, 140–147 (1995).

91. T. Colomb, P. Dahlgren, D. Beghuin, E. Cuche, P. Marquet, and C. Depeursinge, "Polarization imaging by use of digital holography," Applied optics **41**, 27–37 (2002).

92. R. Oldenbourg, "Polarization microscopy with the LC-PolScope," Live Cell Imaging: A Laboratory Manual 205–237 (2005).

93. Z. Wang, L. J. Millet, M. U. Gillette, and G. Popescu, "Jones phase microscopy of transparent and anisotropic samples," Optics letters **33**, 1270–1272 (2008).

94. Y. Kim, J. Jeong, J. Jang, M. W. Kim, and Y. Park, "Polarization holographic microscopy for extracting spatio-temporally resolved Jones matrix," Optics Express **20**, 9948–9955 (2012).

95. S. Aknoun, P. Bon, J. Savatier, B. Wattellier, and S. Monneret, "Quantitative retardance imaging of biological samples using quadriwave lateral shearing interferometry," Optics express **23**, 16383–16406 (2015).

96. S. Aknoun, M. Aurrand-Lions, B. Wattellier, and S. Monneret, "Quantitative retardance imaging by means of quadri-wave lateral shearing interferometry for label-free fiber imaging in tissues," Optics Communications **422**, 17–27 (2018).

97. B. Bai, H. Wang, T. Liu, Y. Rivenson, J. FitzGerald, and A. Ozcan, "Pathological crystal imaging with single-shot computational polarized light microscopy," Journal of biophotonics **13**, e201960036 (2020).

98. Y. Zhang, S. Y. C. Lee, Y. Zhang, D. Furst, J. Fitzgerald, and A. Ozcan, "Wide-field imaging of birefringent synovial fluid crystals using lens-free polarized microscopy for gout diagnosis," Scientific reports **6**, 28793 (2016).

99. C. Oh, S. O. Isikman, B. Khademhosseinieh, and A. Ozcan, "On-chip differential interference contrast microscopy using lensless digital holography," Opt. Express, OE **18**, 4717–4726 (2010).

100. A. Greenbaum, W. Luo, T.-W. Su, Z. Göröcs, L. Xue, S. O. Isikman, A. F. Coskun, O. Mudanyali, and A. Ozcan, "Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy," Nature Methods **9**, 889–895 (2012).

101. W. Bishara, T.-W. Su, A. F. Coskun, and A. Ozcan, "Lensfree on-chip microscopy over a wide field-

of-view using pixel super-resolution," Opt. Express, OE **18**, 11181–11191 (2010).

102. O. Mudanyali, D. Tseng, C. Oh, S. O. Isikman, I. Sencan, W. Bishara, C. Oztoprak, S. Seo, B. Khademhosseini, and A. Ozcan, "Compact, Light-weight and Cost-effective Microscope based on Lensless Incoherent Holography for Telemedicine Applications," Lab Chip **10**, 1417–1428 (2010).

103. A. Ozcan and E. McLeod, "Lensless Imaging and Sensing," Annual Review of Biomedical Engineering **18**, 77–102 (2016).

104. H. Zhu, S. O. Isikman, O. Mudanyali, A. Greenbaum, and A. Ozcan, "Optical imaging techniques for point-of-care diagnostics," Lab Chip **13**, 51–67 (2012).

105. K. de Haan, Y. Rivenson, Y. Wu, and A. Ozcan, "Deep-Learning-Based Image Reconstruction and Enhancement in Optical Microscopy," Proceedings of the IEEE **108**, 30–50 (2020).

106. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica, OPTICA **6**, 921–943 (2019).

107. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica, OPTICA **4**, 1117–1125 (2017).

108. Y. Wu, Y. Rivenson, Y. Zhang, Z. Wei, H. Günaydin, X. Lin, and A. Ozcan, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," Optica **5**, 704–710 (2018).

109. T. Liu, Z. Wei, Y. Rivenson, K. de Haan, Y. Zhang, Y. Wu, and A. Ozcan, "Deep learning-based color holographic microscopy," Journal of biophotonics e201900107 (2019).

110. 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 **8**, 23 (2019).

111. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images," Cell **173**, 792-803.e19 (2018).

112. Y. Intaravanne and X. Chen, "Recent advances in optical metasurfaces for polarization detection and engineered polarization profiles," Nanophotonics **9**, 1003–1014 (2020).

113. R. C. Jones, "A New Calculus for the Treatment of Optical SystemsI. Description and Discussion of

the Calculus," J. Opt. Soc. Am., JOSA **31**, 488–493 (1941).

114. "Register Multimodal MRI Images - MATLAB & Simulink Example," https://www.mathworks.com/help/images/registering-multimodal-mri-images.html.

115. Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Günaydın, Y. Zhang, Z. Gorocs, K. Liang, and D. Tseng, "Deep learning enhanced mobile-phone microscopy," Acs Photonics **5**, 2354–2364 (2018).

116. A. Greenbaum, Y. Zhang, A. Feizi, P.-L. Chung, W. Luo, S. R. Kandukuri, and A. Ozcan, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," Science translational medicine **6**, 267ra175-267ra175 (2014).

117. W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," Light Sci Appl **5**, e16060 (2016).

118. E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," Appl. Opt., AO **38**, 6994–7001 (1999).

119. N. T. Shaked, M. T. Rinehart, and A. Wax, "Dual-interference-channel quantitative-phase microscopy of live cell dynamics," Opt. Lett., OL **34**, 767–769 (2009).

120. L. Tian and L. Waller, "Quantitative differential phase contrast imaging in an LED array microscope," Opt. Express, OE **23**, 11394–11403 (2015).

121. Z. Wang, G. Popescu, K. V. Tangella, and A. Balla, "Tissue refractive index as marker of disease," JBO **16**, 116017 (2011).

122. Z. Wang, H. Ding, and G. Popescu, "Scattering-phase theorem," Opt. Lett., OL **36**, 1215–1217 (2011).

123. Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting Cancer Metastases on Gigapixel Pathology Images," (2017).

124. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis **42**, 60–88 (2017).

125. Y. K. Tao, D. Shen, Y. Sheikine, O. O. Ahsen, H. H. Wang, D. B. Schmolze, N. B. Johnson, J. S. Brooker, A. E. Cable, J. L. Connolly, and J. G. Fujimoto, "Assessment of breast pathologies using

nonlinear microscopy," Proceedings of the National Academy of Sciences **111**, 15304–15309 (2014).

126. M. G. Giacomelli, L. Husvogt, H. Vardeh, B. E. Faulkner-Jones, J. Hornegger, J. L. Connolly, and J. G. Fujimoto, "Virtual Hematoxylin and Eosin Transillumination Microscopy Using Epi-Fluorescence Imaging," PLOS ONE **11**, e0159337 (2016).

127. D. A. Orringer, B. Pandian, Y. S. Niknafs, T. C. Hollon, J. Boyle, S. Lewis, M. Garrard, S. L. Hervey-Jumper, H. J. L. Garton, C. O. Maher, J. A. Heth, O. Sagher, D. A. Wilkinson, M. Snuderl, S. Venneti, S. H. Ramkissoon, K. A. McFadden, A. Fisher-Hubbard, A. P. Lieberman, T. D. Johnson, X. S. Xie, J. K. Trautman, C. W. Freudiger, and S. Camelo-Piragua, "Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy," Nat Biomed Eng **1**, 1–13 (2017).

128. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117 (2017).

129. Y. Jo, H. Cho, S. Y. Lee, G. Choi, G. Kim, H. Min, and Y. Park, "Quantitative Phase Imaging and Artificial Intelligence: A Review," (2018).

130. G. Choi, D. Ryu, Y. Jo, Y. S. Kim, W. Park, H. Min, and Y. Park, "Cycle-consistent deep learning approach to coherent noise reduction in optical diffraction tomography," Opt. Express, OE **27**, 4927–4943 (2019).

131. T. Nguyen, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, "Deep learning approach to Fourier ptychographic microscopy," Opt. Express **26**, 26470 (2018).

132. N. Boyd, E. Jonas, H. Babcock, and B. Recht, "DeepLoco: Fast 3D Localization Microscopy Using Neural Networks," 267096 (2018).

133. T. Zahavy, A. Dikopoltsev, D. Moss, G. I. Haham, O. Cohen, S. Mannor, and M. Segev, "Deep learning reconstruction of ultrashort pulses," Optica, OPTICA **5**, 666–673 (2018).

134. H. S. Park, M. T. Rinehart, K. A. Walzer, J.-T. A. Chi, and A. Wax, "Automated Detection of P. falciparum Using Machine Learning Algorithms with Quantitative Phase Images of Unstained Cells," PLOS ONE **11**, e0163045 (2016).

135. C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali, "Deep Learning in Label-free Cell Classification," Sci Rep **6**, 21471 (2016).

136. D. Roitshtain, L. Wolbromsky, E. Bal, H. Greenspan, L. L. Satterwhite, and N. T. Shaked,

"Quantitative phase microscopy spatial signatures of cancer cells," Cytometry Part A **91**, 482–493 (2017).

137. Y. Jo, S. Park, J. Jung, J. Yoon, H. Joo, M. Kim, S.-J. Kang, M. C. Choi, S. Y. Lee, and Y. Park, "Holographic deep learning for rapid optical screening of anthrax spores," 109108 (2017).
138. B. Javidi, A. Markman, S. Rawat, T. O'Connor, A. Anand, and B. Andemariam, "Sickle cell disease diagnosis based on spatio-temporal cell dynamics analysis using 3D printed shearing digital holographic microscopy," Opt. Express, OE **26**, 13614–13627 (2018).
139. A. Tata, A. Gribble, M. Ventura, M. Ganguly, E. Bluemke, H. J. Ginsberg, D. A. Jaffray, D. R. Ifa, A. Vitkin, and A. Zarrine-Afsar, "Wide-field tissue polarimetry allows efficient localized mass spectrometry imaging of biological tissues," Chem. Sci. **7**, 2162–2169 (2016).
140. I. A. Cree, Z. Deans, M. J. L. Ligtenberg, N. Normanno, A. Edsjö, E. Rouleau, F. Solé, E. Thunnissen, W. Timens, E. Schuuring, E. Dequeker, S. Murray, M. Dietel, P. Groenen, and J. H. V. Krieken, "Guidance for laboratories performing molecular pathology for cancer patients," Journal of Clinical Pathology **67**, 923–931 (2014).
141. P. G. Patel, S. Selvarajah, S. Boursalie, N. E. How, J. Ejdelman, K.-P. Guerard, J. M. Bartlett, J. Lapointe, P. C. Park, J. B. A. Okello, and D. M. Berman, "Preparation of Formalin-fixed Paraffin-embedded Tissue Cores for both RNA and DNA Extraction," JoVE (Journal of Visualized Experiments) e54299 (2016).
142. T. Ikeda, G. Popescu, R. R. Dasari, and M. S. Feld, "Hilbert phase microscopy for investigating fast dynamics in transparent systems," Opt. Lett., OL **30**, 1165–1167 (2005).
143. N. T. Shaked, Y. Zhu, N. Badie, N. Bursac, and A. P. Wax, "Reflective interferometric chamber for quantitative phase imaging of biological sample dynamics," JBO **15**, 030503 (2010).
144. E. Watanabe, T. Hoshiba, and B. Javidi, "High-precision microscopic phase imaging without phase unwrapping for cancer cell identification," Opt. Lett., OL **38**, 1319–1321 (2013).
145. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multiframe Super Resolution," IEEE Transactions on Image Processing **13**, 1327–1344 (2004).
146. J. Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8**, 679–698 (1986).
147. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image

segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.

148. L. Gibellini, A. Cossarizza, M. Pinti, S. De Biasi, and M. Nasi, "Rare Cells: Focus on Detection and Clinical Relevance," in (2017).

149. A. Jindal, P. Gupta, Jayadeva, and D. Sengupta, "Discovery of rare cells from voluminous single cell expression data," Nature Communications **9**, 4719 (2018).

150. E. Arvaniti and M. Claassen, "Sensitive detection of rare disease-associated cell subsets via representation learning," Nature Communications **8**, 14825 (2017).

151. M. Rezaei, M. Winter, D. Zander-Fox, C. Whitehead, J. Liebelt, M. E. Warkiani, T. Hardy, and B. Thierry, "A Reappraisal of Circulating Fetal Cell Noninvasive Prenatal Testing.," Trends in biotechnology **0**, (2018).

152. P. Bacher and A. Scheffold, "Flow-cytometric analysis of rare antigen-specific T cells," Cytometry Part A **83A**, 692–701 (2013).

153. F. Bertolini, Y. Shaked, P. Mancuso, and R. S. Kerbel, "The multifaceted circulating endothelial cell in cancer: towards marker and target identification," Nature Reviews Cancer **6**, 835–845 (2006).

154. J. M. Lang, B. P. Casavant, D. J. Beebe, B. L. Emmink, D. T. Miyamoto, E. Brachtel, M. Yu, P. Chen, B. Morgan, J. Trautwein, A. Kimura, S. Sengupta, S. L. Stott, N. M. Karabacak, T. A. Barber, J. R. Walsh, K. Smith, P. S. Spuhler, J. P. Sullivan, R. J. Lee, D. T. Ting, X. Luo, A. T. Shaw, A. Bardia, L. V. Sequist, D. N. Louis, S. Maheswaran, R. Kapur, D. A. Haber, and M. Toner, "Circulating tumor cells: getting more from less.," Science translational medicine **4**, 141ps13 (2012).

155. S. Massberg, P. Schaerli, I. Knezevic-Maramica, M. Köllnberger, N. Tubo, E. A. Moseman, I. V. Huff, T. Junt, A. J. Wagers, I. B. Mazo, and U. H. von Andrian, "Immunosurveillance by Hematopoietic Progenitor Cells Trafficking through Blood, Lymph, and Peripheral Tissues," Cell **131**, 994–1008 (2007).

156. U. Dharmasiri, M. A. Witek, A. A. Adams, and S. A. Soper, "Microsystems for the Capture of Low-Abundance Cells," Annual Review of Analytical Chemistry **3**, 409–431 (2010).

157. M. Zborowski and J. J. Chalmers, "Rare Cell Separation and Analysis by Magnetic Sorting," Analytical Chemistry **83**, 8050–8056 (2011).

158. Y. Chen, P. Li, P.-H. Huang, Y. Xie, J. D. Mai, L. Wang, N.-T. Nguyen, and T. J. Huang, "Rare cell

isolation and analysis in microfluidics," Lab on a Chip **14**, 626 (2014).

159. Z. Shen, A. Wu, and X. Chen, "Current detection technologies for circulating tumor cells," Chemical Society Reviews **46**, 2038–2056 (2017).

160. F. Zheng, Y. Cheng, J. Wang, J. Lu, B. Zhang, Y. Zhao, and Z. Gu, "Aptamer-Functionalized Barcode Particles for the Capture and Detection of Multiple Types of Circulating Tumor Cells," Advanced Materials **26**, 7333–7338 (2014).

161. A. H. Talasaz, A. A. Powell, D. E. Huber, J. G. Berbee, K.-H. Roh, W. Yu, W. Xiao, M. M. Davis, R. F. Pease, M. N. Mindrinos, S. S. Jeffrey, and R. W. Davis, "Isolating highly enriched populations of circulating epithelial cells and other rare cells from blood using a magnetic sweeper device," Proceedings of the National Academy of Sciences **106**, 3970–3975 (2009).

162. S. Balasubramanian, D. Kagan, C.-M. Jack Hu, S. Campuzano, M. J. Lobo-Castañon, N. Lim, D. Y. Kang, M. Zimmerman, L. Zhang, and J. Wang, "Micromachine-Enabled Capture and Isolation of Cancer Cells in Complex Media," Angewandte Chemie International Edition **50**, 4161–4164 (2011).

163. S. Nagrath, L. V. Sequist, S. Maheswaran, D. W. Bell, D. Irimia, L. Ulkus, M. R. Smith, E. L. Kwak, S. Digumarthy, A. Muzikansky, P. Ryan, U. J. Balis, R. G. Tompkins, D. A. Haber, and M. Toner, "Isolation of rare circulating tumour cells in cancer patients by microchip technology," Nature **450**, 1235–1239 (2007).

164. J.-M. Park, J.-Y. Lee, J.-G. Lee, H. Jeong, J.-M. Oh, Y. J. Kim, D. Park, M. S. Kim, H. J. Lee, J. H. Oh, S. S. Lee, W.-Y. Lee, and N. Huh, "Highly Efficient Assay of Circulating Tumor Cells by Selective Sedimentation with a Density Gradient Medium and Microfiltration from Whole Blood," Analytical Chemistry **84**, 7400–7407 (2012).

165. K. Xiong, W. Wei, Y. Jin, S. Wang, D. Zhao, S. Wang, X. Gao, C. Qiao, H. Yue, G. Ma, and H.-Y. Xie, "Biomimetic Immuno-Magnetosomes for High-Performance Enrichment of Circulating Tumor Cells," Advanced Materials **28**, 7929–7935 (2016).

166. N.-N. Lu, M. Xie, J. Wang, S.-W. Lv, J.-S. Yi, W.-G. Dong, and W.-H. Huang, "Biotin-Triggered Decomposable Immunomagnetic Beads for Capture and Release of Circulating Tumor Cells," ACS Applied Materials & Interfaces **7**, 8817–8826 (2015).

167. S.-I. Han and K.-H. Han, "Electrical Detection Method for Circulating Tumor Cells Using Graphene Nanoplates," Analytical Chemistry **87**, 10585–10592 (2015).

168. M. Y. Sha, H. Xu, M. J. Natan, and R. Cromer, "Surface-Enhanced Raman Scattering Tags for Rapid and Homogeneous Detection of Circulating Tumor Cells in the Presence of Human Whole Blood," Journal of the American Chemical Society **130**, 17214–17215 (2008).

169. L. Wang, P. Balasubramanian, A. P. Chen, S. Kummar, Y. A. Evrard, and R. J. Kinders, "Promise and limits of the CellSearch platform for evaluating pharmacodynamics in circulating tumor cells," Seminars in Oncology **43**, 464–475 (2016).

170. D. Weller, P. Vedsted, G. Rubin, F. M. Walter, J. Emery, S. Scott, C. Campbell, R. S. Andersen, W. Hamilton, F. Olesen, P. Rose, S. Nafees, E. van Rijswijk, S. Hiom, C. Muth, M. Beyer, and R. D. Neal, "The Aarhus statement: improving design and reporting of studies on early cancer diagnosis," British Journal of Cancer **106**, 1262–1267 (2012).

171. R. Etzioni, N. Urban, S. Ramsey, M. McIntosh, S. Schwartz, B. Reid, J. Radich, G. Anderson, and L. Hartwell, "The case for early detection," Nature Reviews Cancer **3**, 243–252 (2003).

172. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437 (2017).

173. Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Günaydın, Y. Zhang, Z. Göröcs, K. Liang, D. Tseng, and A. Ozcan, "Deep Learning Enhanced Mobile-Phone Microscopy," ACS Photonics **5**, 2354–2364 (2018).

174. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Science & Applications **7**, 17141 (2018).

175. Y. Wu, Y. Rivenson, Y. Zhang, Z. Wei, H. Günaydin, X. Lin, and A. Ozcan, "Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery," Optica **5**, 704 (2018).

176. T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," Scientific Reports **9**, (2019).

177. 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 **8**, (2019).

178. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T.

Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering (2019).

179. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nature Methods **16**, 103–110 (2019).

180. Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. de Haan, and A. Ozcan, "Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," Light: Science & Applications **8**, (2019).

181. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods **15**, 1090–1097 (2018).

182. R. Strack, "AI transforms image reconstruction," Nature Methods **15**, 309–309 (2018).

183. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181 (2018).

184. B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," Nature **555**, 487–492 (2018).

185. Z. Göröcs, M. Tamamitsu, V. Bianco, P. Wolf, S. Roy, K. Shindo, K. Yanny, Y. Wu, H. C. Koydemir, Y. Rivenson, and A. Ozcan, "A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples," Light: Science & Applications **7**, (2018).

186. Y. Wu, A. Calis, Y. Luo, C. Chen, M. Lutton, Y. Rivenson, X. Lin, H. C. Koydemir, Y. Zhang, H. Wang, Z. Göröcs, and A. Ozcan, "Label-Free Bioaerosol Sensing Using Mobile Microscopy and Deep Learning," ACS Photonics **5**, 4617–4627 (2018).

187. Y. Zhang, H. Ceylan Koydemir, M. M. Shimogawa, S. Yalcin, A. Guziak, T. Liu, I. Oguz, Y. Huang, B. Bai, Y. Luo, Y. Luo, Z. Wei, H. Wang, V. Bianco, B. Zhang, R. Nadkarni, K. Hill, and A. Ozcan, "Motility-based label-free detection of parasites in bodily fluids using holographic speckle analysis and deep learning," Light: Science & Applications **7**, (2018).

188. G. Kim, D. Ahn, M. Kang, Y. Jo, D. Ryu, H. Kim, J. Song, J. S. Ryu, G. Choi, H. J. Chung, K. Kim,

D. R. Chung, I. Y. Yoo, H. J. Huh, H. Min, N. Y. Lee, and Y. Park, "Rapid and label-free identification of individual bacterial pathogens exploiting three-dimensional quantitative phase imaging and deep learning," bioRxiv (2019).

189. P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," Radiology **284**, 574–582 (2017).

190. B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. A. W. M. van der Laak, and the CAMELYON16 Consortium, M. Hermsen, Q. F. Manson, M. Balkenhol, O. Geessink, N. Stathonikos, M. C. van Dijk, P. Bult, F. Beca, A. H. Beck, D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. Zhong, Q. Dou, Q. Li, H. Chen, H.-J. Lin, P.-A. Heng, C. Haß, E. Bruni, Q. Wong, U. Halici, M. Ü. Öner, R. Cetin-Atalay, M. Berseth, V. Khvatkov, A. Vylegzhanin, O. Kraus, M. Shaban, N. Rajpoot, R. Awan, K. Sirinukunwattana, T. Qaiser, Y.-W. Tsang, D. Tellez, J. Annuscheit, P. Hufnagl, M. Valkonen, K. Kartasalo, L. Latonen, P. Ruusuvuori, K. Liimatainen, S. Albarqouni, B. Mungal, A. George, S. Demirci, N. Navab, S. Watanabe, S. Seno, Y. Takenaka, H. Matsuda, H. Ahmady Phoulady, V. Kovalev, A. Kalinovsky, V. Liauchuk, G. Bueno, M. M. Fernandez-Carrobles, I. Serrano, O. Deniz, D. Racoceanu, and R. Venâncio, "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer," JAMA **318**, 2199 (2017).

191. R. Lindsey, A. Daluiski, S. Chopra, A. Lachapelle, M. Mozer, S. Sicular, D. Hanel, M. Gardner, A. Gupta, R. Hotchkiss, and H. Potter, "Deep neural network improves fracture detection by clinicians," Proceedings of the National Academy of Sciences **115**, 11591–11596 (2018).

192. J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, G. van den Driessche, B. Lakshminarayanan, C. Meyer, F. Mackinder, S. Bouton, K. Ayoub, R. Chopra, D. King, A. Karthikesalingam, C. O. Hughes, R. Raine, J. Hughes, D. A. Sim, C. Egan, A. Tufail, H. Montgomery, D. Hassabis, G. Rees, T. Back, P. T. Khaw, M. Suleyman, J. Cornebise, P. A. Keane, and O. Ronneberger, "Clinically applicable deep learning for diagnosis and referral in retinal disease," Nature Medicine **24**, 1342–1350 (2018).

193. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine **25**, 44–56 (2019).

194. J. N. Anker, C. Behrend, and R. Kopelman, "Aspherical magnetically modulated optical nanoprobes

(MagMOONs)," Journal of Applied Physics **93**, 6698–6700 (2003).

195. J. N. Anker, C. J. Behrend, B. H. McNaughton, T. G. Roberts, M. Brasuel, M. A. Philbert, and R. Kopelman, "Characterization and Applications of Modulated Optical Nanoprobes (MOONs)," MRS Proceedings **790**, (2003).

196. A. Greenbaum, Y. Zhang, A. Feizi, P.-L. Chung, W. Luo, S. R. Kandukuri, and A. Ozcan, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," Science Translational Medicine **6**, 267ra175-267ra175 (2014).

197. A. Feizi, Y. Zhang, A. Greenbaum, A. Guziak, M. Luong, R. Y. L. Chan, B. Berg, H. Ozkan, W. Luo, M. Wu, Y. Wu, and A. Ozcan, "Rapid, portable and cost-effective yeast cell viability and concentration analysis using lensfree on-chip microscopy and machine learning," Lab Chip **16**, 4350–4358 (2016).

198. T.-W. Su, L. Xue, and A. Ozcan, "High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories," Proceedings of the National Academy of Sciences **109**, 16018–16022 (2012).

199. Y. Zhang, S. Y. C. Lee, Y. Zhang, D. Furst, J. Fitzgerald, and A. Ozcan, "Wide-field imaging of birefringent synovial fluid crystals using lens-free polarized microscopy for gout diagnosis," Scientific Reports **6**, 28793 (2016).

200. Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3d residual networks," in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 5533–5541.

201. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).

202. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 4700–4708.

203. H. Ye, Z. Wu, R.-W. Zhao, X. Wang, Y.-G. Jiang, and X. Xue, "Evaluating Two-Stream CNN for Video Classification," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval - ICMR '15* (ACM Press, 2015), pp. 435–442.

204. M. Tamamitsu, Y. Zhang, H. Wang, Y. Wu, and A. Ozcan, "Comparison of Gini index and Tamura coefficient for holographic autofocusing based on the edge sparsity of the complex optical

wavefront," arXiv preprint arXiv:1708.08055 (2017).

205. K. Hajian-Tilaki, "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation," Caspian J Intern Med **4**, 627–635 (2013).

206. S. Krause, M. V. Maffini, A. M. Soto, and C. Sonnenschein, "The microenvironment determines the breast cancer cells' phenotype: organization of MCF7 cells in 3D cultures," BMC Cancer **10**, (2010).

207. N. Fatsis-Kavalopoulos, P. O'Callaghan, B. Xie, R. Hernández Vera, O. Idevall-Hagren, and J. Kreuger, "Formation of precisely composed cancer cell clusters using a cell assembly generator (CAGE) for studying paracrine signaling at single-cell resolution," Lab on a Chip **19**, 1071–1081 (2019).

208. M. Zhao, P. G. Schiro, J. S. Kuo, K. M. Koehler, D. E. Sabath, V. Popov, Q. Feng, and D. T. Chiu, "An Automated High-Throughput Counting Method for Screening Circulating Tumor Cells in Peripheral Blood," Analytical Chemistry **85**, 2465–2471 (2013).

209. X. Wu, Y. Xia, Y. Huang, J. Li, H. Ruan, T. Chen, L. Luo, Z. Shen, and A. Wu, "Improved SERS-Active Nanoparticles with Various Shapes for CTC Detection without Enrichment Process with Supersensitivity and High Specificity," ACS Applied Materials & Interfaces **8**, 19928–19938 (2016).

210. J. Balsam, H. A. Bruck, and A. Rasooly, "Cell streak imaging cytometry for rare cell detection," Biosensors and Bioelectronics **64**, 154–160 (2015).

211. D. Issadore, J. Chung, H. Shao, M. Liong, A. A. Ghazani, C. M. Castro, R. Weissleder, and H. Lee, "Ultrasensitive Clinical Enumeration of Rare Cells ex Vivo Using a Micro-Hall Detector," Science Translational Medicine **4**, 141ra92-141ra92 (2012).

212. W. Liu, H. Wei, Z. Lin, S. Mao, and J.-M. Lin, "Rare cell chemiluminescence detection based on aptamer-specific capture in microfluidic channels," Biosensors and Bioelectronics **28**, 438–442 (2011).

213. T. Gao, L. Li, T. Chen, L. Shi, Y. Yang, and G. Li, "DNA-Oriented Shaping of Cell Features for the Detection of Rare Disseminated Tumor Cells," Analytical Chemistry **91**, 1126–1132 (2019).

214. B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," IEEE Transactions on Image Processing **5**, 1266–1271 (1996).

215. D. Jiles, *Introduction to Magnetism and Magnetic Materials*, 2. ed (Chapman & Hall, 1998).

216. C. J. Murray, K. S. Ikuta, F. Sharara, L. Swetschinski, G. R. Aguilar, A. Gray, C. Han, C. Bisignano,

P. Rao, E. Wool, S. C. Johnson, A. J. Browne, M. G. Chipeta, F. Fell, S. Hackett, G. Haines-Woodhouse, B. H. K. Hamadani, E. A. P. Kumaran, B. McManigal, R. Agarwal, S. Akech, S. Albertson, J. Amuasi, J. Andrews, A. Aravkin, E. Ashley, F. Bailey, S. Baker, B. Basnyat, A. Bekker, R. Bender, A. Bethou, J. Bielicki, S. Boonkasidecha, J. Bukosia, C. Carvalheiro, C. Castañeda-Orjuela, V. Chansamouth, S. Chaurasia, S. Chiurchiù, F. Chowdhury, A. J. Cook, B. Cooper, T. R. Cressey, E. Criollo-Mora, M. Cunningham, S. Darboe, N. P. J. Day, M. D. Luca, K. Dokova, A. Dramowski, S. J. Dunachie, T. Eckmanns, D. Eibach, A. Emami, N. Feasey, N. Fisher-Pearson, K. Forrest, D. Garrett, P. Gastmeier, A. Z. Giref, R. C. Greer, V. Gupta, S. Haller, A. Haselbeck, S. I. Hay, M. Holm, S. Hopkins, K. C. Iregbu, J. Jacobs, D. Jarovsky, F. Javanmardi, M. Khorana, N. Kissoon, E. Kobeissi, T. Kostyanev, F. Krapp, R. Krumkamp, A. Kumar, H. H. Kyu, C. Lim, D. Limmathurotsakul, M. J. Loftus, M. Lunn, J. Ma, N. Mturi, T. Munera-Huertas, P. Musicha, M. M. Mussi-Pinhata, T. Nakamura, R. Nanavati, S. Nangia, P. Newton, C. Ngoun, A. Novotney, D. Nwakanma, C. W. Obiero, A. Olivas-Martinez, P. Olliaro, E. Ooko, E. Ortiz-Brizuela, A. Y. Peleg, C. Perrone, N. Plakkal, A. Ponce-de-Leon, M. Raad, T. Ramdin, A. Riddell, T. Roberts, J. V. Robotham, A. Roca, K. E. Rudd, N. Russell, J. Schnall, J. A. G. Scott, M. Shivamallappa, J. Sifuentes-Osornio, N. Steenkeste, A. J. Stewardson, T. Stoeva, N. Tasak, A. Thaiprakong, G. Thwaites, C. Turner, P. Turner, H. R. van Doorn, S. Velaphi, A. Vongpradith, H. Vu, T. Walsh, S. Waner, T. Wangrangsimakul, T. Wozniak, P. Zheng, B. Sartorius, A. D. Lopez, A. Stergachis, C. Moore, C. Dolecek, and M. Naghavi, "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis," The Lancet **399**, 629–655 (2022).

217. G. Maki and M. Zervos, "Health Care–Acquired Infections in Low- and Middle-Income Countries and the Role of Infection Prevention and Control," Infect Dis Clin North Am **35**, 827–839 (2021).

218. CDC, "CDC Partners Estimate Healthcare Cost of AR Infections," https://www.cdc.gov/drugresistance/solutions-initiative/stories/partnership-estimates-healthcare-cost.html.

219. Environmental Protection Agency, Office of Water, "US Environmental Protection Agency. Method 1604: Total Coliforms and Escherichia Coli in Water by Membrane Filtration Using A Simultaneous Detection Technique (MI Medium).," (2002).

220. US Environmental Protection Agency, "US EPA. Analytical Methods Approved for Compliance

Monitoring under the Long Term 2 Enhanced Surface Water Treatment Rule," (2017).

221. P. Belgrader, W. Benett, D. Hadley, J. Richards, P. Stratton, R. Mariella, and F. Milanovich, "PCR Detection of Bacteria in Seven Minutes," Science **284**, 449–450 (1999).

222. A. J. Blaschke, C. Heyrend, C. L. Byington, M. A. Fisher, E. Barker, N. F. Garrone, S. A. Thatcher, A. T. Pavia, T. Barney, G. D. Alger, J. A. Daly, K. M. Ririe, I. Ota, and M. A. Poritz, "Rapid identification of pathogens from positive blood cultures by multiplex polymerase chain reaction using the FilmArray system," Diagnostic Microbiology and Infectious Disease **74**, 349–355 (2012).

223. V. A. Kempf, K. Trebesius, and I. B. Autenrieth, "Fluorescent In situ hybridization allows rapid identification of microorganisms in blood cultures," J Clin Microbiol **38**, 830–838 (2000).

224. H. Iseoka, M. Sasai, S. Miyagawa, K. Takekita, S. Date, H. Ayame, A. Nishida, S. Sanami, T. Hayakawa, and Y. Sawa, "Rapid and sensitive mycoplasma detection system using image-based deep learning," J Artif Organs **25**, 50–58 (2022).

225. H. Wang, H. Ceylan Koydemir, Y. Qiu, B. Bai, Y. Zhang, Y. Jin, S. Tok, E. C. Yilmaz, E. Gumustekin, Y. Rivenson, and A. Ozcan, "Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning," Light Sci Appl **9**, 118 (2020).

226. A. Pangajam, K. Theyagarajan, and K. Dinakaran, "Highly sensitive electrochemical detection of E. coli O157:H7 using conductive carbon dot/ZnO nanorod/PANI composite electrode," Sensing and Bio-Sensing Research **29**, 100317 (2020).

227. A. Ramanujam, B. Neyhouse, R. A. Keogh, M. Muthuvel, R. K. Carroll, and G. G. Botte, "Rapid electrochemical detection of Escherichia coli using nickel oxidation reaction on a rotating disk electrode," Chem Eng J **411**, 128453 (2021).

228. S. N. A. Jenie, Y. Kusumastuti, F. S. H. Krismastuti, Y. M. Untoro, R. T. Dewi, L. Z. Udin, and N. Artanti, "Rapid Fluorescence Quenching Detection of Escherichia coli Using Natural Silica-Based Nanoparticles," Sensors (Basel) **21**, 881 (2021).

229. A. D. Taylor, Q. Yu, S. Chen, J. Homola, and S. Jiang, "Comparison of E. coli O157:H7 preparation methods used for detection with surface plasmon resonance sensor," Sensors and Actuators B: Chemical **107**, 202–208 (2005).

230. J. Chen, A. A. Jackson, V. M. Rotello, and S. R. Nugen, "Colorimetric Detection of Escherichia coli Based on the Enzyme-Induced Metallization of Gold Nanorods," Small **12**, 2469–2475 (2016).

231. L. Song, L. Zhang, Y. Huang, L. Chen, G. Zhang, Z. Shen, J. Zhang, Z. Xiao, and T. Chen, "Amplifying the signal of localized surface plasmon resonance sensing for the sensitive detection of Escherichia coli O157:H7," Sci Rep **7**, 3288 (2017).

232. M. R. Nurliyana, M. Z. Sahdan, K. M. Wibowo, A. Muslihati, H. Saim, S. A. Ahmad, Y. Sari, and Z. Mansor, "The Detection Method of Escherichia coli in Water Resources: A Review," J. Phys.: Conf. Ser. **995**, 012065 (2018).

233. I. Tryland, H. Braathen, A. C. Wennberg, F. Eregno, and A.-L. Beschorner, "Monitoring of β-d-Galactosidase Activity as a Surrogate Parameter for Rapid Detection of Sewage Contamination in Urban Recreational Water," Water **8**, 65 (2016).

234. S. o. Van Poucke and H. j. Nelis, "A 210-min solid phase cytometry test for the enumeration of Escherichia coli in drinking water," Journal of Applied Microbiology **89**, 390–396 (2000).

235. P. P. Banada, S. Guo, B. Bayraktar, E. Bae, B. Rajwa, J. P. Robinson, E. D. Hirleman, and A. K. Bhunia, "Optical forward-scattering for detection of Listeria monocytogenes and other Listeria species," Biosensors and Bioelectronics **22**, 1664–1671 (2007).

236. A. Suchwalko, I. Buzalewicz, A. Wieliczko, and H. Podbielska, "Bacteria species identification by the statistical analysis of bacterial colonies Fresnel patterns," Opt. Express, OE **21**, 11322–11337 (2013).

237. R. London, J. Schwedock, A. Sage, H. Valley, J. Meadows, M. Waddington, and D. Straus, "An Automated System for Rapid Non-Destructive Enumeration of Growing Microbes," PLOS ONE **5**, e8609 (2010).

238. C.-S. Ho, N. Jean, C. A. Hogan, L. Blackmon, S. S. Jeffrey, M. Holodniy, N. Banaei, A. A. E. Saleh, S. Ermon, and J. Dionne, "Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning," Nat Commun **10**, 4927 (2019).

239. H. Zhu, U. Sikora, and A. Ozcan, "Quantum dot enabled detection of Escherichia coli using a cell-phone," Analyst **137**, 2541–2544 (2012).

240. V. Müller, J. M. Sousa, H. C. Koydemir, M. Veli, D. Tseng, L. Cerqueira, A. Ozcan, N. F. Azevedo, and F. Westerlund, "Identification of pathogenic bacteria in complex samples using a smartphone based fluorescence microscope," RSC Adv. **8**, 36493–36502 (2018).

241. S. Tok, K. de Haan, D. Tseng, C. F. Usanmaz, H. Ceylan Koydemir, and A. Ozcan, "Early detection

of E. coli and total coliform using an automated, colorimetric and fluorometric fiber optics-based device," Lab Chip **19**, 2925–2935 (2019).

242. M. Mizukami, S. Oku, S.-I. Cho, M. Tatetsu, M. Abiko, M. Mamada, T. Sakanoue, Y. Suzuri, J. Kido, and S. Tokito, "A Solution-Processed Organic Thin-Film Transistor Backplane for Flexible Multiphoton Emission Organic Light-Emitting Diode Displays," IEEE Electron Device Letters **36**, 841–843 (2015).

243. V. Fiore, P. Battiato, S. Abdinia, S. Jacobs, I. Chartier, R. Coppard, G. Klink, E. Cantatore, E. Ragonese, and G. Palmisano, "An Integrated 13.56-MHz RFID Tag in a Printed Organic Complementary TFT Technology on Flexible Substrate," IEEE Transactions on Circuits and Systems I: Regular Papers **62**, 1668–1677 (2015).

244. M. Kaltenbrunner, T. Sekitani, J. Reeder, T. Yokota, K. Kuribara, T. Tokuhara, M. Drack, R. Schwödiauer, I. Graz, S. Bauer-Gogonea, S. Bauer, and T. Someya, "An ultra-lightweight design for imperceptible plastic electronics," Nature **499**, 458–463 (2013).

245. Y. Takeda, K. Hayasaka, R. Shiwaku, K. Yokosawa, T. Shiba, M. Mamada, D. Kumaki, K. Fukuda, and S. Tokito, "Fabrication of Ultra-Thin Printed Organic TFT CMOS Logic Circuits Optimized for Low-Voltage Wearable Sensor Applications," Sci Rep **6**, 25714 (2016).

246. A. Kumar, A. K. Goyal, and N. Gupta, "Review—Thin-Film Transistors (TFTs) for Highly Sensitive Biosensing Applications: A Review," ECS J. Solid State Sci. Technol. **9**, 115022 (2020).

247. Y. S. Rim, "Review of metal oxide semiconductors-based thin-film transistors for point-of-care sensor applications," Journal of Information Display **21**, 203–210 (2020).

248. C. Hu, S. Kalsi, I. Zeimpekis, K. Sun, P. Ashburn, C. Turner, J. M. Sutton, and H. Morgan, "Ultra-fast electronic detection of antimicrobial resistance genes using isothermal amplification and Thin Film Transistor sensors," Biosensors and Bioelectronics **96**, 281–287 (2017).

249. R. A. S. Domínguez, O. O. Hinostroza, A. O. Díaz, and M. Á. D. Jiménez, "Antibody Immobilization in ZnO-Thin Film Transistors for Low-Cost Biosensors Applications," Engineering Proceedings **4**, 23 (2021).

250. R. A. Salinas, A. Orduña-Díaz, O. Obregon-Hinostroza, and M. A. Dominguez, "Biosensors based on zinc oxide thin-film transistors using recyclable plastic substrates as an alternative for real-time pathogen detection," Talanta **237**, 122970 (2022).

251. R. A. Street, W. S. Wong, T. Ng, and R. Lujan, "Amorphous silicon thin film transistor image sensors," Philosophical Magazine **89**, 2687–2697 (2009).

252. Y. Duan, G. Coatrieux, and H. Shu, "Identification of digital radiography image source based on digital radiography pattern noise recognition," in *2014 IEEE International Conference on Image Processing (ICIP)* (2014), pp. 5372–5376.

253. N. Lu, W. Jiang, Q. Wu, D. Geng, L. Li, and M. Liu, "A Review for Compact Model of Thin-Film Transistors (TFTs)," Micromachines **9**, 599 (2018).

254. D. Tseng, O. Mudanyali, C. Oztoprak, S. O. Isikman, I. Sencan, O. Yaglidere, and A. Ozcan, "Lensfree microscopy on a cellphone," Lab Chip **10**, 1787–1792 (2010).

255. O. Mudanyali, D. Tseng, C. Oh, S. O. Isikman, I. Sencan, W. Bishara, C. Oztoprak, S. Seo, B. Khademhosseini, and A. Ozcan, "Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications," Lab Chip **10**, 1417–1428 (2010).

256. A. Greenbaum, Y. Zhang, A. Feizi, P.-L. Chung, W. Luo, S. R. Kandukuri, and A. Ozcan, "Wide-field computational imaging of pathology slides using lens-free on-chip microscopy," Science Translational Medicine **6**, 267ra175-267ra175 (2014).

257. Z. Göröcs and A. Ozcan, "On-Chip Biomedical Imaging," IEEE Rev Biomed Eng **6**, 29–46 (2013).

258. Y. Wu and A. Ozcan, "Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring," Methods **136**, 4–16 (2018).

259. N. Malpica, C. O. de Solórzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. García-Sagredo, and F. del Pozo, "Applying watershed algorithms to the segmentation of clustered nuclei," Cytometry **28**, 289–297 (1997).

260. G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, and K. Weinberger, "Convolutional Networks with Dense Connectivity," IEEE Transactions on Pattern Analysis and Machine Intelligence 1–1 (2019).

261. Z. Qiu, T. Yao, and T. Mei, "Learning Spatio-Temporal Representation With Pseudo-3D Residual Networks," in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 5533–5541.

262. L. Singh, H. G. Kruger, G. E. M. Maguire, T. Govender, and R. Parboosing, "The role of nanotechnology in the treatment of viral infections," Ther Adv Infect Dis **4**, 105–131 (2017).

263. CDC, "Burden of Influenza," https://www.cdc.gov/flu/about/burden/index.html.

264. CDC, "Preliminary In-Season 2021-2022 Flu Burden Estimates," https://www.cdc.gov/flu/about/burden/preliminary-in-season-estimates.htm.

265. "COVID-19 Map," https://coronavirus.jhu.edu/map.html.

266. W.-S. Ryu, "Diagnosis and Methods," Molecular Virology of Human Pathogenic Viruses 47–62 (2017).

267. Z. Wen, M. Citron, A. J. Bett, A. S. Espeseth, K. A. Vora, L. Zhang, and D. J. DiStefano, "Development and application of a higher throughput RSV plaque assay by immunofluorescent imaging," Journal of Virological Methods **263**, 88–95 (2019).

268. S. Basak, H.-J. Kang, K.-B. Chu, J. Oh, and F.-S. Quan, "Simple and rapid plaque assay for recombinant baculoviruses expressing influenza hemagglutinin," Science Progress **104**, 00368504211004261 (2021).

269. M. Abou-Karam and W. T. Shier, "A Simplified Plaque Reduction Assay for Antiviral Agents from Plants. Demonstration of Frequent Occurrence of Antiviral Activity in Higher Plants," J. Nat. Prod. **53**, 340–344 (1990).

270. P. C. Stepp and Ph.D., "New Method for Rapid Virus Quantification," GEN - Genetic Engineering and Biotechnology News (2010).

271. W.-S. Ryu, *Molecular Virology of Human Pathogenic Viruses* (Academic Press, an imprint of Elsevier, 2017).

272. E. J. Mendoza, K. Manguiat, H. Wood, and M. Drebot, "Two Detailed Plaque Assay Protocols for the Quantification of Infectious SARS-CoV-2," Current Protocols in Microbiology **57**, cpmc105 (2020).

273. K. A. Rashid, S. Hevi, Y. Chen, F. L. Cahérec, and S. L. Chuck, "A Proteomic Approach Identifies Proteins in Hepatocytes That Bind Nascent Apolipoprotein B *," Journal of Biological Chemistry **277**, 22010–22017 (2002).

274. J. A. Blaho, E. R. Morton, and J. C. Yedowitz, "Herpes Simplex Virus: Propagation, Quantification, and Storage," Current Protocols in Microbiology **00**, 14E.1.1-14E.1.23 (2006).

275. D. J. M. Cruz and H.-J. Shin, "Application of a focus formation assay for detection and titration of porcine epidemic diarrhea virus," J Virol Methods **145**, 56–61 (2007).

276. S. Loret, N. El Bilali, and R. Lippé, "Analysis of herpes simplex virus type I nuclear particles by

flow cytometry," Cytometry Part A **81A**, 950–959 (2012).

277. S. D. Gallaher and A. J. Berk, "A rapid Q-PCR titration protocol for adenovirus and helper-dependent adenovirus vectors that produces biologically relevant results," J Virol Methods **192**, 28–38 (2013).

278. M. L. Killian, "Hemagglutination Assay for Influenza Virus," in *Animal Influenza Virus*, E. Spackman, ed., Methods in Molecular Biology (Springer, 2014), pp. 3–9.

279. P. Roingeard, P. Raynal, S. Eymieux, and E. Blanchard, "Virus detection by transmission electron microscopy: Still useful for diagnosis and a plus for biosafety," Rev Med Virol **29**, e2019 (2019).

280. M. Alhajj and A. Farhana, "Enzyme Linked Immunosorbent Assay," in *StatPearls* (StatPearls Publishing, 2022).

281. K. Pankaj, "Virus Identification and Quantification," Materials and Methods (2021).

282. A. Baer and K. Kehn-Hall, "Viral Concentration Determination Through Plaque Assays: Using Traditional and Novel Overlay Systems," J Vis Exp 52065 (2014).

283. A. L. Masci, E. B. Menesale, W.-C. Chen, C. Co, X. Lu, and S. Bergelson, "Integration of Fluorescence Detection and Image-Based Automated Counting Increases Speed, Sensitivity, and Robustness of Plaque Assays," Mol Ther Methods Clin Dev **14**, 270–274 (2019).

284. N. Ke, X. Wang, X. Xu, and Y. A. Abassi, "The xCELLigence System for Real-Time and Label-Free Monitoring of Cell Viability," in *Mammalian Cell Viability: Methods and Protocols*, M. J. Stoddart, ed., Methods in Molecular Biology (Humana Press, 2011), pp. 33–43.

285. G. Burmakina, K. Bliznetsov, and A. Malogolovkin, "Real-time analysis of the cytopathic effect of African swine fever virus," Journal of Virological Methods **257**, 58–61 (2018).

286. Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," Nature Photon **12**, 578–589 (2018).

287. T. Cacace, V. Bianco, and P. Ferraro, "Quantitative phase imaging trends in biomedical applications," Optics and Lasers in Engineering **135**, 106188 (2020).

288. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light Sci Appl **7**, 17141–17141 (2018).

289. Y. Luo, L. Huang, Y. Rivenson, and A. Ozcan, "Single-Shot Autofocusing of Microscopy Images Using Deep Learning," ACS Photonics **8**, 625–638 (2021).

290. H. Ding, F. Li, Z. Meng, S. Feng, J. Ma, S. Nie, and C. Yuan, "Auto-focusing and quantitative phase

imaging using deep learning for the incoherent illumination microscopy system," Opt. Express, OE **29**, 26385–26403 (2021).

291. T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," Sci Rep **9**, 3926 (2019).

292. G. Kim, D. Ahn, M. Kang, J. Park, D. Ryu, Y. Jo, J. Song, J. S. Ryu, G. Choi, H. J. Chung, K. Kim, D. R. Chung, I. Y. Yoo, H. J. Huh, H. Min, N. Y. Lee, and Y. Park, "Rapid label-free identification of pathogenic bacteria species from a minute quantity exploiting three-dimensional quantitative phase imaging and artificial neural network," 596486 (2021).

293. Ç. Işıl, K. de Haan, Z. Göröcs, H. C. Koydemir, S. Peterman, D. Baum, F. Song, T. Skandakumar, E. Gumustekin, and A. Ozcan, "Phenotypic Analysis of Microalgae Populations Using Label-Free Imaging Flow Cytometry and Deep Learning," ACS Photonics **8**, 1232–1242 (2021).

294. A. Butola, D. Popova, D. K. Prasad, A. Ahmad, A. Habib, J. C. Tinguely, P. Basnet, G. Acharya, P. Senthilkumaran, D. S. Mehta, and B. S. Ahluwalia, "High spatially sensitive quantitative phase imaging assisted with deep neural network for classification of human spermatozoa under stressed condition," Sci Rep **10**, 13118 (2020).

295. G. Kim, Y. Jo, H. Cho, H. Min, and Y. Park, "Learning-based screening of hematologic disorders using quantitative phase imaging of individual red blood cells," Biosensors and Bioelectronics **123**, 69–76 (2019).

296. X. Shu, S. Sansare, D. Jin, X. Zeng, K.-Y. Tong, R. Pandey, and R. Zhou, "Artificial-Intelligence-Enabled Reagent-Free Imaging Hematology Analyzer," Advanced Intelligent Systems **3**, 2000277 (2021).

297. T. O'Connor, A. Anand, B. Andemariam, and B. Javidi, "Deep learning-based cell identification and disease diagnosis using spatio-temporal cellular dynamics in compact digital holographic microscopy," Biomed. Opt. Express, BOE **11**, 4491–4508 (2020).

298. N. Goswami, Y. R. He, Y.-H. Deng, C. Oh, N. Sobh, E. Valera, R. Bashir, N. Ismail, H. Kong, T. H. Nguyen, C. Best-Popescu, and G. Popescu, "Label-free SARS-CoV-2 detection and classification using phase imaging with computational specificity," Light Sci Appl **10**, 176 (2021).

299. T. O'Connor, J.-B. Shen, B. T. Liang, and B. Javidi, "Digital holographic deep learning of red blood cells for field-portable, rapid COVID-19 screening," Opt. Lett., OL **46**, 2344–2347 (2021).

300. Y. Li, T. Liu, H. C. Koydemir, H. Wang, K. O’Riordan, B. Bai, Y. Haga, J. Kobashi, H. Tanaka, T. Tamaru, K. Yamaguchi, and A. Ozcan, *Deep Learning-Enabled Detection and Classification of Bacterial Colonies Using a Thin Film Transistor (TFT) Image Sensor* (arXiv, 2022).

301. Y. Zhang, H. Ceylan Koydemir, M. M. Shimogawa, S. Yalcin, A. Guziak, T. Liu, I. Oguz, Y. Huang, B. Bai, Y. Luo, Y. Luo, Z. Wei, H. Wang, V. Bianco, B. Zhang, R. Nadkarni, K. Hill, and A. Ozcan, "Motility-based label-free detection of parasites in bodily fluids using holographic speckle analysis and deep learning," Light Sci Appl **7**, 108 (2018).

302. W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," Light Sci Appl **5**, e16060–e16060 (2016).

303. A. A. Abdelmageed and M. C. Ferran, "The Propagation, Quantification, and Storage of Vesicular Stomatitis Virus," Curr Protoc Microbiol **58**, e110 (2020).

304. M. Tamamitsu, Y. Zhang, H. Wang, Y. Wu, and A. Ozcan, "A robust holographic autofocusing criterion based on edge sparsity: comparison of Gini index and Tamura coefficient for holographic autofocusing based on the edge sparsity of the complex optical wavefront," in *Quantitative Phase Imaging IV* (SPIE, 2018), Vol. 10503, pp. 22–31.

305. "Estimate displacement field that aligns two 2-D or 3-D images - MATLAB imregdemons," https://www.mathworks.com/help/images/ref/imregdemons.html.

306. G. Reitberger and T. Sauer, "Background Subtraction using Adaptive Singular Value Decomposition," J Math Imaging Vis **62**, 1159–1172 (2020).