Title
Deep Learning-Enabled Computational Imaging in Optical Microscopy and Air Quality Monitoring

Author
Wu, Yichen

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Deep Learning-Enabled Computational Imaging in Optical Microscopy and Air Quality Monitoring

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

by

Yichen Wu

2019

ABSTRACT OF THE DISSERTATION

Deep Learning-Enabled Computational Imaging

in Optical Microscopy and Air Quality Monitoring

by

Yichen Wu

Doctor of Philosophy in Electrical and Computer Engineering

University of California, Los Angeles, 2019

Professor Aydogan Ozcan, Chair

Exponential advancements in computational resources and algorithms have given birth to the new paradigm in imaging that rely on computation to digitally reconstruct and enhance images. These computational imaging modalities have enabled higher resolution, larger throughput and/or automatic detection capabilities for optical microscopy. An example is lens-less digital holographic microscope, which enables snapshot imaging of volumetric samples over wide field-of-view without using imaging lenses. Recent developments in the field of deep learning have further opened up exciting avenues for computational imaging, which offer unprecedented performance thanks to their capability to robustly learn content-specific complex image priors.

This dissertation introduces a novel and universal modeling framework of deep learning - based image reconstruction technique to tackle various challenges in optical microscopic

imaging, including digital holography reconstructions and 3D fluorescence microscopy. Firstly, auto-focusing and phase recovery in holography reconstruction are conventionally challenging and time-consuming to digitally perform. A convolutional neural network (CNN) based approach was developed that solves both problems rapidly in parallel, enabling extended depth-of-field holographic reconstruction with significantly improved time complexity from *O(mn)* to *O(1)*. Secondly, to fuse advantages of snapshot volumetric capability in digital holography and speckle- and artifact-free image contrast in bright-field microscopy, a CNN was used to transform *across microscopy modalities* from holographic image reconstructions to their equivalent high contrast bright-field microscopic images. Thirdly, 3D fluorescence microscopy generally requires axial scanning. A CNN was trained to learn defocuses of fluorescence and digitally refocusing a single 2D fluorescence image onto user-defined 3D surfaces within the sample volume, which extends depth-of-field of fluorescence microscopy by *20-fold without any axial scanning, additional hardware, or a trade-off of imaging resolution or speed*. This enables high-speed volumetric imaging and digital aberration correction for live samples.

Based on deep learning powered computational microscopy, a hand-held device was also developed to measure the particulate matters and bio-aerosols in the air using the lens-less digital holographic microscopic imaging geometry. This device, named *c-Air*, demonstrates accurate, high-throughput and automatic detection, sizing and classification of the particles in the air, which opens new opportunities in deep learning based environmental sensing and personalized and/or distributed air quality monitoring.

The dissertation of Yichen Wu is approved.

Benjamin Williams

Oscar M. Stafsudd

Mona Jarrahi

Aydogan Ozcan, Committee Chair

University of California, Los Angeles

2019

Table of Contents

Acknowledgements

I would like to express my sincere gratitude to Prof. Aydogan Ozcan. I feel very fortunate to work with Prof. Ozcan and be part of his flourishing lab. His unique vision to the field, his open-mindedness to new ideas, and his insightful guidance have all been essential to my research and professional development. Prof. Ozcan has also been a personal role model for me. His unparalleled dedication to innovation and excellence, as well as his brilliant communication skills, has inspired me a lot. In addition, I would like to thank my doctoral committee members, Prof. Mona Jarrahi, Prof. Benjamin Williams and Prof. Oscar Stafsudd for their kind support during my Ph.D.

I would like to acknowledge my colleges: Steve Feng, Ashutosh Shiledar, Jeffrey Wong Yicheng Li, and Christine Chen for helping to establish some of my earliest Ph.D. work; Dr. Yair Rivenson, Dr. Ani Ray, Prof. Yifang Zhu, Dr. Yibo Zhang, Dr. Zoltán Göröcs and Dr. Wei Luo for their informal guidance of my researches through collaborations and discussions; Yi Luo, Yilin Luo, Gunvant Chaudhari, Xin Tong, Dr. Ayfer Calis, and Dr. Laurent Bentolila for their recent but very valuable contributions; as well as Derek Tseng, Kyle Liang, Zach Ballard, Dr. Eyal Ben-David, Dr. Michael Lake, Dr. Euan Mcleod, Alborz Feizi, Dr. Qingshan Wei, Dr. Hatice Ceylan Koydemir, Hongda Wang, Xuan Cheng, Maxwell Lutton, Serene Kamal, Cheng Chen, among many others, for their kind help during my Ph.D.

Last but not least, I would like to thank my family for their unconditional support. They have always been by my side to cheer me up when I met bottlenecks and cool me down when I felt too satisfied.

Vita

Yichen Wu joined the Bio- and Nano-photonics Lab led by Prof. Aydogan Ozcan at UCLA in Fall 2014. He received his Bachelor of Science in Engineering (B.S.E.) in Information Engineering (Optics) from Zhejiang University, China in 2014. He has co-authored 21 journal articles, and various conference proceedings and patents on computational and bio-medical imaging and sensing. He is a winner of UCLA Dissertation Year Fellowship, SPIE John Kiel scholarship, Vodafone Americas Wireless Innovation Award, and UCLA EE departmental fellowship. Yichen also served as the 2017-2018 president of the UCLA Chapter of OSA&SPIE.

Selected Publications

1. **Wu, Yichen**, Yair Rivenson, Hongda Wang, Yilin Luo, Eyal Ben-David, and Aydogan Ozcan. "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning." Nature Methods 2019 (accepted).
2. Yair Rivenson, **Yichen Wu**, and Aydogan Ozcan. "Deep Learning in Holography and Coherent Imaging." Light: Science & Applications 8, no. 1 (September 11, 2019): 1–8.
3. Rivenson, Yair, Hongda Wang, Zhensong Wei, Kevin de Haan, Yibo Zhang, **Yichen Wu**, Harun Günaydın, et al. "Virtual Histological Staining of Unlabelled Tissue-Autofluorescence Images via Deep Learning." Nature Biomedical Engineering, March 4, 2019, 1.
4. **Wu, Yichen**, Yilin Luo, Gunvant Chaudhari, Yair Rivenson, Ayfer Calis, Kevin de Haan, and Aydogan Ozcan. "Bright-Field Holography: Cross-Modality Deep Learning Enables Snapshot 3D Imaging with Bright-Field Contrast Using a Single Hologram." Light: Science & Applications 8, no. 1 (March 6, 2019): 25.
5. Haan, Kevin de, Zachary Ballard, Yair Rivenson, **Yichen Wu**, and Aydogan Ozcan. "Resolution Enhancement in Scanning Electron Microscopy Using Deep Learning." Scientific Reports 9, no. 1 (August 19, 2019): 1–7.
6. **Yichen Wu**, Aniruddha Ray, Qingshan Wei, Alborz Feizi, Xin Tong, Eva Chen, Yi Luo, and Aydogan Ozcan. "Deep Learning Enables High-Throughput Analysis of Particle-Aggregation-Based Biosensors Imaged Using Holography." ACS Photonics, December 21, 2018.
7. **Wu, Yichen**, Ayfer Calis, Yi Luo, Cheng Chen, Maxwell Lutton, Yair Rivenson, Xing Lin, et al. "Label-Free

Bio-Aerosol Sensing Using Mobile Microscopy and Deep Learning." ACS Photonics, October 4, 2018.

8. Göröcs, Zoltán, Miu Tamamitsu, Vittorio Bianco, Patrick Wolf, Shounak Roy, Koyoshi Shindo, Kyrollos Yanny, **Yichen Wu**, Hatice Ceylan Koydemir, Yair Rivenson, Aydogan 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, no. 1 (September 19, 2018): 66.
9. Zhang, Yibo, Tairan Liu, Yujia Huang, Da Teng, Yinxu Bian, **Yichen Wu**, Yair Rivenson, Alborz Feizi, and Aydogan Ozcan. "Accurate Color Imaging of Pathology Slides Using Holography and Absorbance Spectrum Estimation of Histochemical Stains." Journal of Biophotonics 0, no. 0 (October 23, 2018): e201800335.
10. **Wu, Yichen**, Yair Rivenson, Yibo Zhang, Zhensong Wei, Harun Günaydin, Xing Lin, and Aydogan Ozcan. "Extended Depth-of-Field in Holographic Imaging Using Deep-Learning-Based Autofocusing and Phase Recovery." Optica 5, no. 6 (June 20, 2018): 704–10.
11. **Wu, Yichen**, and Aydogan Ozcan. "Lensless Digital Holographic Microscopy and Its Applications in Biomedicine and Environmental Monitoring." Methods 136 (March 1, 2018): 4–16.
12. Zhang, Yibo, Hongda Wang, **Yichen Wu**, Miu Tamamitsu, and Aydogan Ozcan. "Edge Sparsity Criterion for Robust Holographic Autofocusing." Optics Letters 42, no. 19 (October 1, 2017): 3824–27.
13. Tamamitsu, Miu, Yibo Zhang, Hongda Wang, **Yichen Wu**, and Aydogan 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.
14. **Wu, Yichen**, Ashutosh Shiledar, Yicheng Li, Jeffrey Wong, Steve Feng, Xuan Chen, Christine Chen, et al. "Air Quality Monitoring Using Mobile Microscopy and Machine Learning." Light: Science & Applications 6, no. 9 (September 2017): e17046–e17046.
15. Rivenson, Yair, **Yichen Wu**, Hongda Wang, Yibo Zhang, Alborz Feizi, and Aydogan Ozcan. "Sparsity-Based Multi-Height Phase Recovery in Holographic Microscopy." Scientific Reports 6 (November 30, 2016): 37862.
16. **Wu, Yichen,** Yibo Zhang, Wei Luo, and Aydogan Ozcan. "Demosaiced Pixel Super-Resolution for Multiplexed Holographic Color Imaging." Scientific Reports 6 (June 29, 2016): 28601.
17. Yibo Zhang, **Yichen Wu**, Yun Zhang, and Aydogan Ozcan. "Color calibration and fusion of lens-free and mobile-phone microscopy images for high-resolution and accurate color reproduction." Scientific reports 6 (2016): 27811.

Chapter 1 Computational microscopy and digital holography

For centuries, biomedical imaging has been powered by optical compound microscopes, leading to numerous discoveries at the micro- and nano-scale [1,2]. Over the last few decades, as the computational resources become exponentially faster, cheaper, more powerful and portable [3], computational microscopy have emerged that rely on computation to digitally generate high-resolution images over large sample areas and volumes [4–7]. Among these different computational microscopy techniques, digital holography is one of the most widely explored modalities as it permits high-throughput 3D imaging of phase and amplitude information of specimen [8–15].

A specific example is lens-less digital holographic on-chip microscopy [16,17], which digitally reconstructs microscopic images of specimens without using any lenses, as a result of which it can be made much smaller, lighter and lower-cost (**Figure 1.1(a)**). Furthermore, the limited space-bandwidth product of objective lenses in a conventional microscope can be significantly surpassed by a lens-less microscope, enabling ultra-wide field of view (FOV) that are orders of magnitude larger than the FOV of a conventional microscope (see **Figure 1.1(b)** for a comparison). Such lens-less imaging designs have enabled high-resolution and high-throughput imaging of specimens using compact, portable and cost-effective devices to potentially address various point-of-care, global-health and telemedicine related challenges, including cancer and disease diagnosis [4,18,19], environmental monitoring [20–24], microbial viability testing [25], 3D motion tracking of biological samples [26–29], analyzing high-energy particle tracks [30], among others [31], [23].

In this chapter, I will first derive a general formulation of imaging system by its point spread function (PSF), which forms the foundation of various computational imaging approaches that will be expanded throughout this dissertation. Next, using the example of lens-less microscope, I will discuss design and operation principles behind digital holography imaging and image

reconstruction, as well as various computational techniques to improve the quality and/or efficiency for image reconstruction. This chapter also includes a brief discussion of my contributions in sparsity based iterative phase recovery [32] and simultaneous super resolution and color imaging for digital holography [33], in sections 1.8 and 1.9 , respectively.

Part of this chapter have previously been published in Y. Wu, and A. Ozcan. "Lensless Digital Holographic Microscopy and Its Applications in Biomedicine and Environmental Monitoring." *Methods*, 136 (March 1, 2018): 4–16.

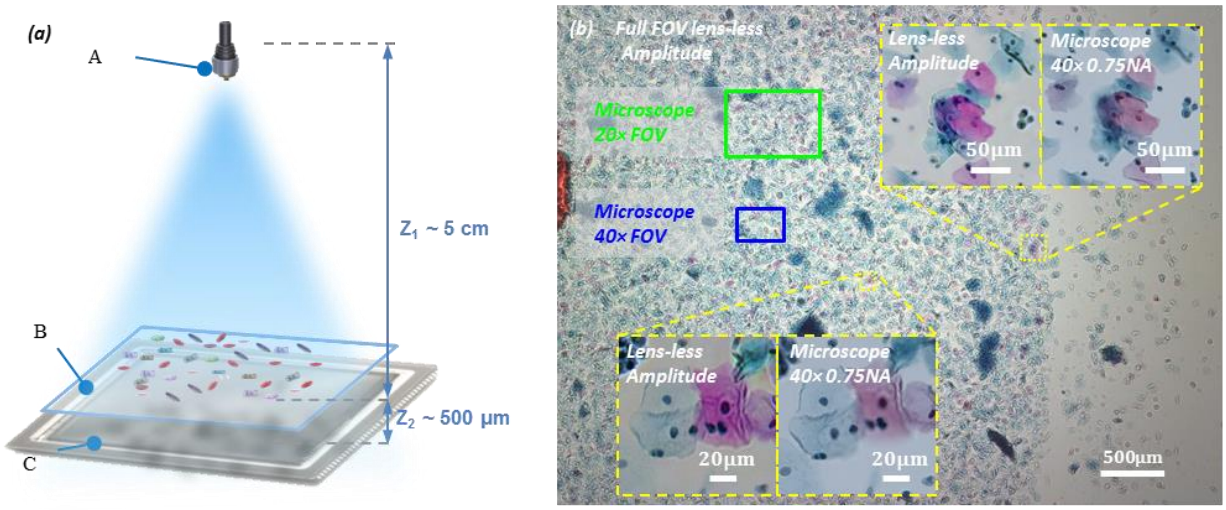

Figure 1.1 Lens-less holographic on-chip microscopy. (a) Schematics of a lens-less on-chip microscope, including (A) a partially coherent light source (e.g., a multi-mode fiber-coupled LED), (B) the sample to be imaged (sample plane), and (C) image sensor chip (sensor plane). (b) Color reconstruction of a Pap smear. The reconstruction resolution and image quality are comparable to conventional microscope images, shown in the inset. Reprinted from Ref. [34]

1.1 Modeling of an imaging systems by PSF and inverse problems

Many imaging systems can be considered as a *linear* system [35]. Specifically, for coherent imaging system such as holography, the system response is linear to the complex amplitude because each point on the object is synchronized; in contrast to the incoherent imaging system

such as bright field microscopy or fluorescence microscopy, where the points on the object is independent to each other and the system response is linear to image intensity.

Linearity of the imaging system leads to the definition of PSF, which can be considered as the impulse response of the imaging system, i.e. for an object function $f(x,y) = \int_{-\infty}^{\infty} f(u,v) \cdot \delta(x-u, y-v)dudv$, its output to the linear imaging system $H\{\cdot\}$ is:

$$g(x,y) = H\{f(x,y)\} = \int_{-\infty}^{\infty} f(u,v) \cdot H\{\delta(x-u, y-v)\}dudv \tag{1.1}$$

where the $h(x,y;u,v) = H\{\delta(x-u, y-v)\}$ is defined as system PSF. In discrete representations, this forward model of a linear imaging system can be equivalently written as:

$$g = Hf + n \tag{1.2}$$

where $f \in R^N$ and $g \in R^M$ are the discrete (vectorized) representation of the object $f(x,y)$ and system output image $g(x,y)$, respectively, $n \in R^M$ is the additive noise, and $H \in R^{M \times N}$ is the discrete matrix representation of the integral with PSF $h(x,y;u,v)$ in equation (1.1). The PSF incorporates different properties of the system, including but not limited to, the finite aperture of microscope objective lens (diffraction limit), size and spatial responsivity of the image sensor pixels, blurring, and system aberrations etc., into a single measurable property. A major task in computational imaging, known as *image restoration (or reconstruction)*, is to solve the inverse problem of equation (1.2), i.e., restore the object function f given one or more measured images g, with or without knowledge of the system function H and noise model n. This task in general is ill-conditioned as the solution is not unique, but some of these inverse problems can be robustly solved using iterative methods, which I will expand in sections 1.6, 1.7 and 1.8; and more recently using non-iterative deep learning approaches, which I will expand in Chapter 2.

For some imaging systems, the PSF can be further assumed to be *spatially-invariant*, i.e., $h(x,y;u,v) = h(x-u,y-v)$, which leads to the *convolution relation*:

$$g(x,y) = \int_{-\infty}^{\infty} f(u,v) \cdot h(x-u,y-v)dudv = f(x,y) \star h(x,y) \tag{1.3}$$

where $\star$ represents convolution operation. In discrete representations, this becomes:

$$g = h \star f + n \tag{1.4}$$

If we take the Fourier transform on both side of equation (1.3), the convolution becomes point-wise multiplication for the system and object function, i.e.,

$$G(f_x, f_y) = F(f_x, f_y) * H(f_x, f_y) \tag{1.5}$$

$$G = H * F + n \tag{1.6}$$

where * represents point-wise multiplication, and G, F, H are Fourier transforms of g, f, h, respectively. This means, the convolution relation of a spatially-invariant PSF facilitates the Fourier analysis of the microscopy system, which not only provides additional insights to the system in the spatial frequency domain, but also enables efficient computation through fast Fourier transforms (FFTs). One example is numerical propagation for digital holography reconstruction, as will be detailed in section 1.5.

1.2 Theory of digital in-line holography

Holography encodes three-dimensional (3D) information of sample through the interference of the scattered sample light and the un-scattered reference light in a hologram, which can be both in-line or off-axis geometries [35]. In in-line holography, the sample wave and reference wave co-propagate in the same direction, whereas there is an angle between the two waves in off-axis holography. In holography, the object is semi-transparent and it can be approximated as:

$$t(x_0, y_0) = 1 + \Delta t(x_0, y_0) \tag{1.7}$$

where $\Delta t \ll 1$. When this object is locally illuminated by a plane wave A, it will propagate *coherently* over a distance of z_2, where z_2 is the spacing between the sample and sensor planes:

$$R[z_2]\{A \cdot t(x_0, y_0)\} = R[z_2]\{A\} + R[z_2]\{A \cdot \Delta t(x_0, y_0)\} = A' + a(x, y) \qquad (1.8)$$

where $R[z_2]\{\cdot\}$ is the free-space propagation operator over a depth of z_2. A hologram is formed by the interference of the scattered sample beam $a(x, y)$ with the un-scattered reference beam A', and the intensity of this interference is recorded by e.g., an image sensor chip:

$$I(x, y) = |A' + a(x, y)|^2 = |A'|^2 + A'^* \cdot a(x, y) + A' \cdot a^*(x, y) + |a(x, y)|^2 \qquad (1.9)$$

where the object-related holographic information is encoded within the second and third terms, which are complex conjugates of each other; the first term can be subtracted out using a background image (without the object present), and the fourth term (i.e., $|a(x, y)|^2$) can be ignored since $\Delta t \ll 1$. In fact, this last term represents the self-interference of the scattered waves from the specimen, which does not contain any useful information as far as holography is concerned. Digital holography digitally reconstructs an image of the object from its hologram. In section 1.5, I will introduce basic reconstruction method for digital holography, which will be followed by more advanced approaches in sections 1.6, 1.7 and 1.8 as well as in Chapter 2.

1.3 Design principles of lens-less microscope

Lens-less on-chip microscope stems from in-line holography (**Figure 1.1(a)**), where a semi-transparent sample is placed on top of an image sensor with a typical spacing of < 1 mm (z_2 distance). A partially coherent light source illuminates at > 2-3 cm (z_1 distance) above the sample; as a result, the sample casts an in-line hologram, recorded by an image sensor. The light source can be either a monochromator [36,37] or a laser [4,38] in a benchtop system or a light emitting diode (LED) in a portable device [17,39] with an optional spectral filter (to fine tune the temporal coherence) [40]. The benefits of a partially coherent source (both temporally and spatially),

compared to its coherent counterpart (e.g., a laser source) include reduced speckle and multiple reflection interference noise as well as much easier automated alignment of digital images for e.g., lateral and axial resolution enhancement[16,41]. Unlike other digital holographic microscopes, an on-chip microscope does not need to use a small (e.g., sub-micron) pinhole since $z_1 >> z_2$; in fact it can operate with very large apertures (e.g., 50-100 μm) without affecting its spatial resolution (detailed below and in Ref. 9). Although the coherence is partial, free space diffraction over >2-3 cm (i.e., z_1 distance) creates enough coherence on the sample/sensor plane so that each individual point can be effectively considered coherent with respect to its local neighbors. In addition, because $z_1 \gg z_2$, the magnification of lens-less microscopy is unit, meaning that the FOV is equal to the entire active area of the image sensor chip.

There are several factors that can limit the resolution of a holographic on-chip microscope, including e.g., diffraction, pixel size, image chip area and coherence of the system. Theoretically, if the pixel size can be arbitrarily small and the coherence is perfect over a large sensor area, then an image that is ideally reconstructed using a lens-less on-chip microscope is *diffraction limited*, with a maximum detectable spatial frequency of $\sim n/\lambda$, where n is the refractive index of the medium between the sample and the sensor plane, and λ is the illumination wavelength.

In practice, an important limitation of the lens-less on-chip microscope resolution is created by the pixel size of the sensor due to its unit magnification. From *Nyquist sampling theorem*, the half-pitch resolution of a reconstructed image equals to the pixel size of the image sensor chip, which using a state-of-art CMOS image sensor chip (e.g., the ones used in smart phone cameras) will be about one micron. With additional image processing, e.g., pixel super-resolution algorithms (section 1.4), an effective pixel size that is much smaller than the wavelength of light can be digitally synthesized from a series of images that are shifted with respect to each other.

The coherence of the illumination can be another source of resolution limitation due to partial spatial coherence and/or temporal coherence. In a lens-less on-chip holographic microscope, the spatial coherence diameter at the sample plane is proportional to $\lambda \cdot z_1/D$, where D is the aperture size at the illumination plane, which can be a simple pinhole or a multi-mode fiber. For a given λ and z_1, the choice of D affects not only the diameter of the spatial coherence at the sensor plane (which determines the effective numerical aperture in an object's in-line hologram) but also the spatial smearing that is induced at the hologram plane. To expand on the latter point, a given illumination aperture, $T(x, y)$, effectively smears the hologram intensity[17,42], i.e.,

$$I^{meas} = I^{coh} \star T\left(-\frac{z_1}{z_2}x, -\frac{z_1}{z_2}y\right) \tag{1.10}$$

where I^{coh} is the hologram assuming perfect spatial coherence over a large image sensor area (e.g., using an infinitesimally small pinhole), and I^{meas} is the measured hologram, and $\star$ stands for the spatial convolution operation. Equation (1.10) states that the aperture function is de-magnified by a factor of $M = z_1/z_2$ and projected onto the hologram plane. Intuitively, this convolution relation of Equation (1.10) can be thought as an incoherent super-position of an infinite number of holograms, each of which is generated by an infinitesimal point source inside the aperture region, and assuming a perfectly incoherent illumination source across $T(x, y)$, each one of these point sources is incoherent with respect to others (**Figure 1.2 (a)**). To avoid this resolution limit due to pinhole or illumination aperture size, the projected/de-magnified aperture needs to be smaller than the pixel size so that the projected aperture equivalently merges into the pixel function. For an on-chip holographic microscope, this is easy to satisfy since $M = \frac{z_1}{z_2} \gg 1$; for example, for an aperture diameter of 100 µm, and a z_2 distance of 400 µm, a z_1 distance of >3-4 cm should be sufficient for a typical CMOS imager chip that has a pixel size of ~1-2 µm.

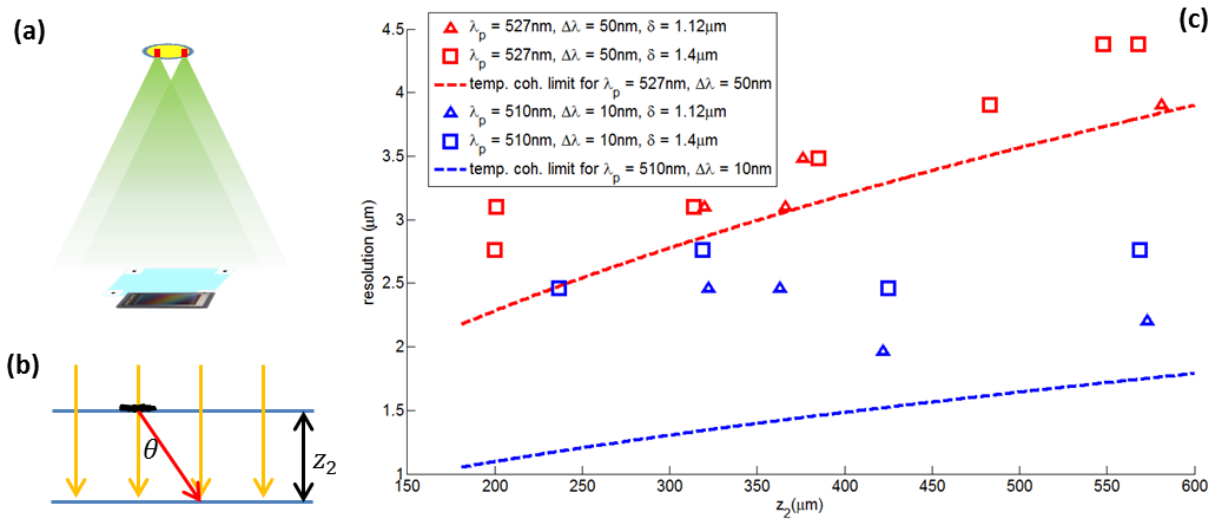

Figure 1.2 Limitation of resolution by coherence in lens-less microscope. (a) Spatial coherence limit and source smearing effect. As a geometric intuition, individual point sources within the aperture generates super-imposed and shifted hologram that adds in intensity, so the spatial coherence limits the resolution in the sense that a scaled aperture convolves with the hologram. (b) Temporal coherence limit. The coherence length of optical wave limits the maximum angle of spatial resolution to interfere and be recorded. (c) Resolution characterization experiment using USAF-1951 resolution target. The illumination is set very far away so spatial coherence doesn't play a role. Temporal coherence plays an important role in resolution when the spectrum bandwidth is not narrow enough (red). With sufficient temporal coherence, the period resolution is around twice the pixel size (blue).

The temporal coherence of the illumination is another factor that is crucial for the spatial resolution of an on-chip holographic microscope and is determined by the coherence length ΔL_c of illumination source, which is a function of the spectral bandwidth $\Delta\lambda$ of the source [43]:

$$\Delta L_c \approx \sqrt{\frac{2\ln 2}{\pi}} \cdot \frac{\lambda^2}{n \cdot \Delta\lambda} \tag{1.11}$$

where n is refractive index, and λ is illumination wavelength. As illustrated in **Figure 1.2(b)**, In order for a spatial frequency component of the object to be recorded at the sensor plane, the difference of the optical path-length of this high frequency scattered wave and the vertical reference wave should not exceed the temporal coherence length. Accordingly, the maximum

angle (θ_{max}) for any scattered plane wave component that can contribute to an object's hologram can be determined as:

$$\cos\theta_{max} = \frac{z_2}{z_2 + \Delta L_c} \tag{1.12}$$

Based on this temporal coherence length consideration, for a z_2 distance of ~ 400 μm, as an example, the bandwidth of the partially coherent light source needs to be smaller than 20 nm in order to achieve a resolution of ~ 1 μm (which is close to the pixel size). **Figure 1.2 (c)** plots the analytical trend (equation (1.12)) and some experimental result of temporal coherence limit on resolution image resolution as a function of z_2 distances, under two spectral bandwidths, center wavelengths and pixel sizes at the image sensor. In this, the z_1 distance is selected to be sufficiently large so that the spatial coherence and the illumination aperture do not compromise resolution. For a narrower illumination bandwidth, and assuming that the spatial coherence does not pose a limitation, then the resolution of an on-chip holographic microscope is limited by the under-sampling that is created by the pixel pitch of the sensor chip, which can be overcome using e.g., pixel super-resolution techniques, as will be detailed in section 1.4.

1.4 Pixel super-resolution for lens-less digital holographic imaging

As discussed in the previous section, one limiting factor for resolution in an on-chip holographic microscope is pixelation and thus under-sampling of spatial frequency information at the hologram plane. To achieve resolution beyond the pixel-pitch limit, we can employ a technique called pixel super-resolution (PSR) [44–46]. In PSR, the object's hologram is shifted laterally in sub-pixel increments, and at each location on the shifting grid, a low resolution (LR) hologram is captured. This relative lateral shift of the hologram with respect to the sensor array plane can be achieved by e.g., shifting the image sensor chip [4], the sample [47], or the illumination source [46,48]. In a portable on-chip imaging device, the mechanical shifts required for PSR can be

replaced by an array of LEDs that are sequentially turned on/off to cast sub-pixel shifted holograms of a static sample [48]. Using multiple LR holograms, a higher-resolution hologram can be digitally synthesized. One algorithm that can be used for this purpose is called "shift-and-add", in which the low-resolution holograms are up-sampled, shifted, and digitally added [49].

The shift-and-add algorithm, although simple in concept and computation, is based on the assumption that the LR holograms are sub-pixel shifted on a *uniform* grid, and that pixel function is ideally a delta function, $\delta(x,y)$. This assumption puts part of the burden on hardware alignment and partially limits the choice of image sensors that can be used. When these assumptions for shift-and-add are harder to satisfy, more versatile techniques, which are based on e.g., iterative optimization, can also be used [45,46]. These iterative methods generally solve an optimization problem to minimize a cost function such as:

$$x^* = \arg\min \sum_i \|W_i \cdot x - y_i\|_p^q + \alpha \cdot \gamma(x) \tag{1.13}$$

This cost function typically contains two parts: the first part $\sum_i \|W_i \cdot x - y_i\|_p^q$ uses some norm (e.g., p=1 or 2 with q=1 or 2, respectively) to minimize the distance between the optimal solution (x^*) and i different measurements, where W_i represents the digital process of shifting and down-sampling of an image and y_i is a lower resolution measurement. The second part $\gamma(x)$ is a regularization parameter to maintain/regulate some desired quality in the reconstructed image, for instance, smoothness (derivative) [45], sparsity (l_1-norm) [50], or sparsity in its derivative (total variation) [44,51], where α serves as a coefficient that balances the strength of regularization term. Since the transformations represented by W_i are linear, this cost function is typically convex [52], and can be optimized via convex optimization algorithms, for instance, gradient-based and conjugate-gradient based descent algorithms [45,49]. Through iterations, some of the spatial artifacts that are typically present in "shift-and-add" based solutions due to the non-uniform shift

grid or finite pixel function, can be minimized. When applied to digital holographic on-chip imaging [46], the high frequency fringes of an object's hologram become quite visible after PSR (see **Figure 1.3(d-e)**), which is used to reconstruct higher resolution images of the samples. By using PSR with a vertical only illumination angle, periodic structures with a line-width as small as ~300 nm was resolved at 530 nm illumination wavelength, equivalent to an effective numerical aperture (NA) of ~0.9 (**Figure 1.3(f)**) [15,16].

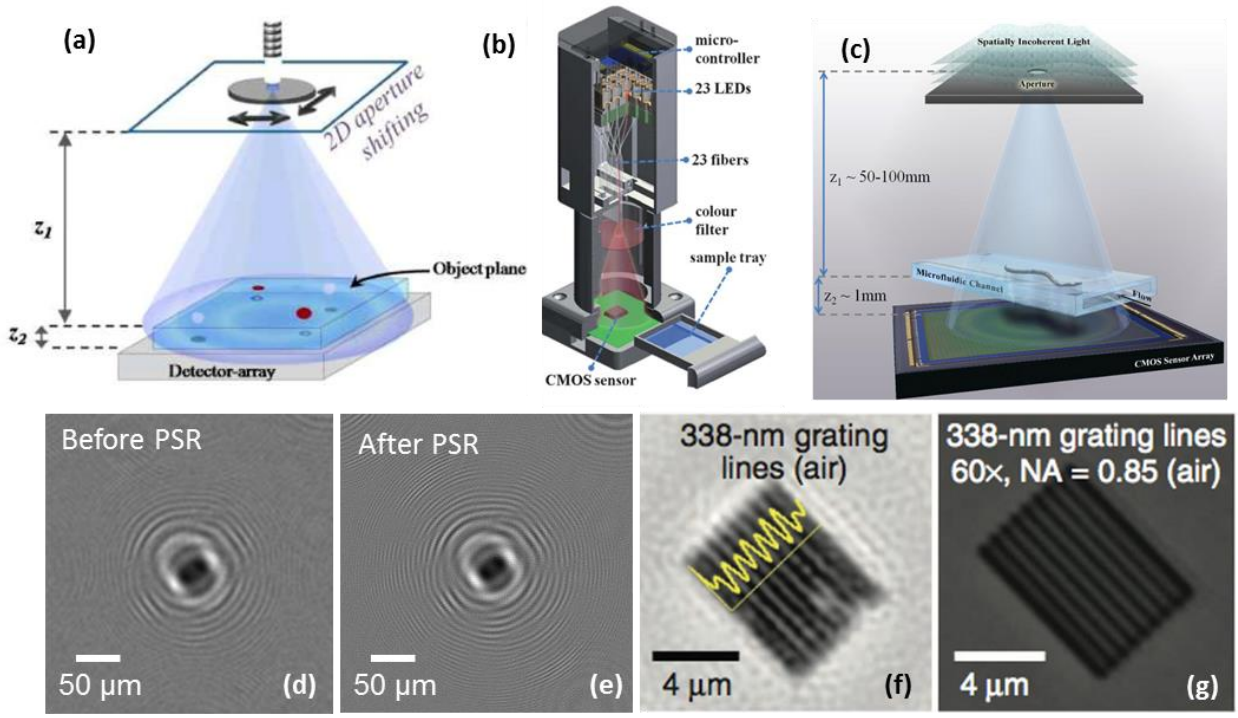

Figure 1.3 Pixel super-resolution (PSR) for digital holographic on-chip imaging. Subpixels shifts in a digital hologram can be achieved through (a) source shifting, (b) an array of static light sources, e.g., fiber-coupled LEDs that are individually controllable, sensor shifting, or (c) sample shifting using e.g., a micro-fluidic flow. (d) An in-line hologram before PSR. (e) Pixel super-resolved version of the same object's hologram. (f) Reconstructed gratings with 338 nm line-width using PSR. (g) Microscope comparison of the same region of interest. Reprinted from Ref. [34].

In addition to sampling sub-pixel shifted holograms through e.g., source, sample or sensor shifting, coherent imaging also offers other methods to achieve super-resolved pixels, for

instance, by varying the illumination wavelengths with small increments (e.g., ~ 2-3 nm) [53], or by varying the propagation distances (i.e., z_2) [54,55]. Some recent work has also used deep-learning to achieve PSR [56–60], also covering microscopy applications [60].

1.5 Digital hologram reconstruction and auto-focusing

In digital holography, the image of a sample, both amplitude and phase, can be reconstructed through a digital back-propagation step. One method of digital back-propagation is the *angular spectrum method* [35]. The angular spectrum is defined in the Fourier domain following the Helmholtz wave equation for the complex amplitude $u_z(x,y)$:

$$\left(\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)u_z(x,y)+k^2u_z(x,y)=0 \tag{1.14}$$

where $k=2\pi n/\lambda$ is the wave-number for illumination wavelength λ and refractive index n of the medium. Taking the Fourier transform of $U(f_x,f_y)=FT\{u(x,y)\}$, we have:

$$\frac{\partial^2}{\partial z^2}U_z(f_x,f_y)+\left[k^2-(2\pi f_x)^2-\left(2\pi f_y\right)^2\right]U_z(f_x,f_y)=0 \tag{1.15}$$

which is a wave equation and has solution (for $k^2\geq(2\pi f_x)^2+\left(2\pi f_y\right)^2$) as:

$$U_z(f_x,f_y)=U_0(f_x,f_y)\cdot e^{j\sqrt{k^2-(2\pi f_x)^2-\left(2\pi f_y\right)^2}\cdot z}=U_0(f_x,f_y)\cdot H(f_x,f_y) \tag{1.16}$$

This means, to numerically propagate a complex wave, the complex wave can be first transformed into spatial frequency (Fourier) domain, multiplied by a propagation kernel $H(f_x,f_y)$ (also known as *angular spectrum*, which is a function of the propagation distance (z)), and is finally transformed back into the spatial domain, using e.g., FFTs. Due to the periodicity of discrete Fourier transform, symmetric padding or tapered average padding on the hologram are frequently used before propagation to avoid *ringing artifacts*. Zero-padding in the frequency (Fourier) domain can also be used to effectively interpolate the propagated image in spatial

domain, which maintains the contrast by avoiding artifacts that are introduced using other image interpolation method such as bicubic interpolation [61].

In lens-less microscope (**Figure 1.1(a)**), the propagation distance equals to the distance between the sample plane and the sensor plane (i.e. $z = z_2$). Accurate estimation of this sample-to-sensor distance (z_2) of different parts of the 3D object is crucial to the quality of the reconstructed holographic image. This z_2 distance estimation step is known as auto-focusing. A widely accepted auto-focusing method involves digitally propagating the recorded hologram to various axial distances, where a focusing criteria is evaluated on each propagated image, and a local maxima or minima of the criteria is selected as the focal plane. Various auto-focusing criteria has been evaluated for digital holographic imaging, including Tamura coefficient [62], edge sparsity [63] and others [64–69]. Edge sparsity has been demonstrated as one of the most robust criteria in auto-focusing, among others [63,70]. To speed up the auto-focusing search process, I also employed a fast searching algorithm, consisting of two steps: a rough focus step to search for a concave interval, and a fine focus step to find the exact focus based on golden ratio search [23].

Since the opto-electronic sensor is only sensitive to the intensity of the impinging light, the back-propagation step converts the object's hologram (equation (1.9)) into four different terms:

$$R[-z_2]\{A^* \cdot I(x,y)\} = |A|^2 \cdot [1 + \Delta t(x,y) + R[-2z_2]\{\Delta t^*(x,y)\}] + R[-z_2]\{|a(x,y)|^2\} \qquad (1.17)$$

where the first term on the right-side is a *DC-background*, the second term is the *reconstructed object* function, the third term is the *twin image artifact* and the fourth term is the back-propagated *self-interference* term. This is related to the fact that the recorded hologram has only amplitude information and its phase is initially missing [71]. In some imaging applications [26,27], where the objects of interest are small and relatively isolated/sparse, the twin image does not pose a significant challenge for the quality of the reconstructed image, and a simple back-propagation can then be sufficient for these applications. In others, the image quality would be

compromised by the twin image and self-interference related noise terms, thus needs to be improved using e.g., *phase recovery* techniques, as will be detailed in subsequent sections.

1.6 Phase recovery using object support

One of the basic phase retrieval methods uses of an object mask constraint [17,71], where a threshold defines a 2D mask on the object plane, and parts of the back-propagated hologram on the object plane that lie outside of the mask regions are considered noise and are iteratively removed. In addition to thresholding, this 2D object mask/support can also be defined through a microscope image of the same region, using e.g., a hybrid design that includes a traditional lens-based microscope as well as a lens-less on-chip microscope [72]. This object-mask constraint works very well for imaging objects that are relatively small and isolated with a clear boundary. As an example, **Figure 1.4** demonstrates the flowchart of a phase recovery algorithm, using object support which is quite effective in eliminating both the twin image noise as well as the multiple interferences in the background (shown as the ripples), enhancing the contrast and signal to noise ratio (SNR) for the objects of interest (isolated particles) in the output image.

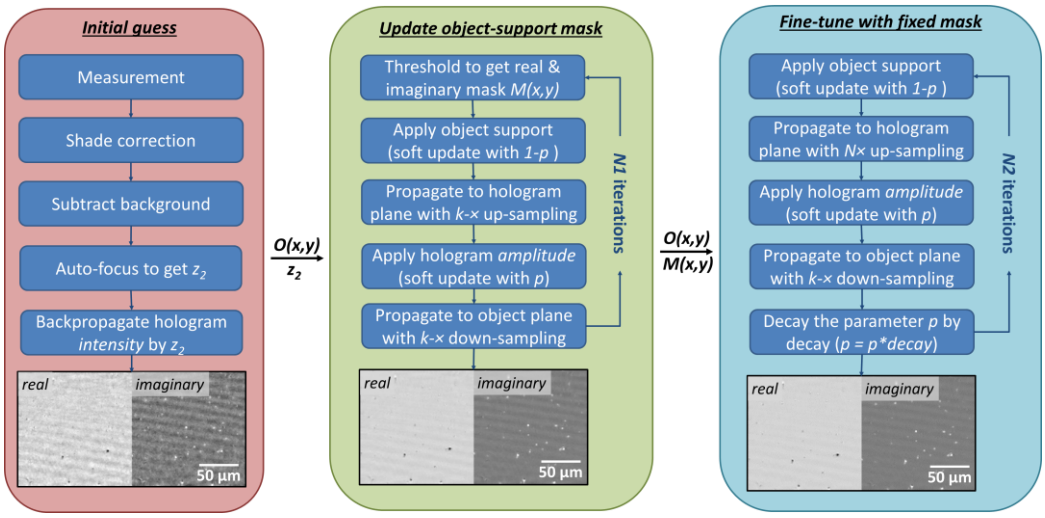

Figure 1.4 Object-support phase recovery for holographic images of aerosol particles. Hyperparameters: p=0.5, decay = 0.95, k = 2, N1 = 100, N2 = 10. Threshold defined as 3 times background standard deviation from background mean. Processing time is ~3 s for a 512×512 patch on GPU.

However, for samples that are confluent and connected, e.g., tissue slides or blood smears, it is more challenging to depict an object mask. More advanced algorithms using additional measurement constraints are needed for phase retrieval in these types of spatially dense and connected samples, which will be detailed in the following sub-sections. Furthermore, as will be discussed in Chapter 2, there are also emerging deep-learning and convolutional neural network based holographic image reconstruction and phase recovery methods that achieve excellent elimination of twin image noise and self-interference related spatial artifacts even for a dense and connected sample and using a single measurement constraint.

1.7 Phase recovery using measurement diversities

For imaging dense and connected samples, phase recovery methods that use additional measurements at different sample-to-sensor distances (heights) can be used. These lens-less holographic measurements at different heights impose physical measurement constraints that the twin image and other sources of noise terms do not satisfy. From these multiple lens-less measurements at different heights, a complex-valued object image can be recovered numerically, by digitally propagating among different measurement planes and replacing the amplitude of the calculated field at each plane by the square root of the measured intensity, while keeping the phase information unchanged. This process can be repeated for e.g., 10-30 iterations until a converged solution is achieved. The general framework of this type of an algorithm is called the Gerchberg-Saxton iterative error-reduction method [71,73]. In different publications, it is also referred to as the iterative projection method [74,75], or multi-height phase recovery algorithm [4,19,76].

In lens-less on-chip imaging, the multi-height phase recovery process typically requires 6-8 heights to efficiently suppress the twin image noise and other spatial artifacts. Because the phase recovery problem is in general not convex, the solution may stagnate at a local optimum, which

is also known as the phase stagnation problem [4,77]. One way to tackle this is to have an initial phase guess that is closer to the global optimum. The transport of intensity equation (TIE) [78,79] has been demonstrated to be effective in providing an initial phase guess for the iterative multi-height phase recovery to enable rapid convergence [4]. Using this framework, lens-less on-chip imaging of pathology slides with a resolution and image quality/contrast comparable to a high-end conventional compound microscope has been demonstrated (**Figure 1.5**) [4,80]. In another study, a field-portable and cost-effective lens-less on-chip microscope based on the same multi-height phase retrieval method was also demonstrated [40,81].

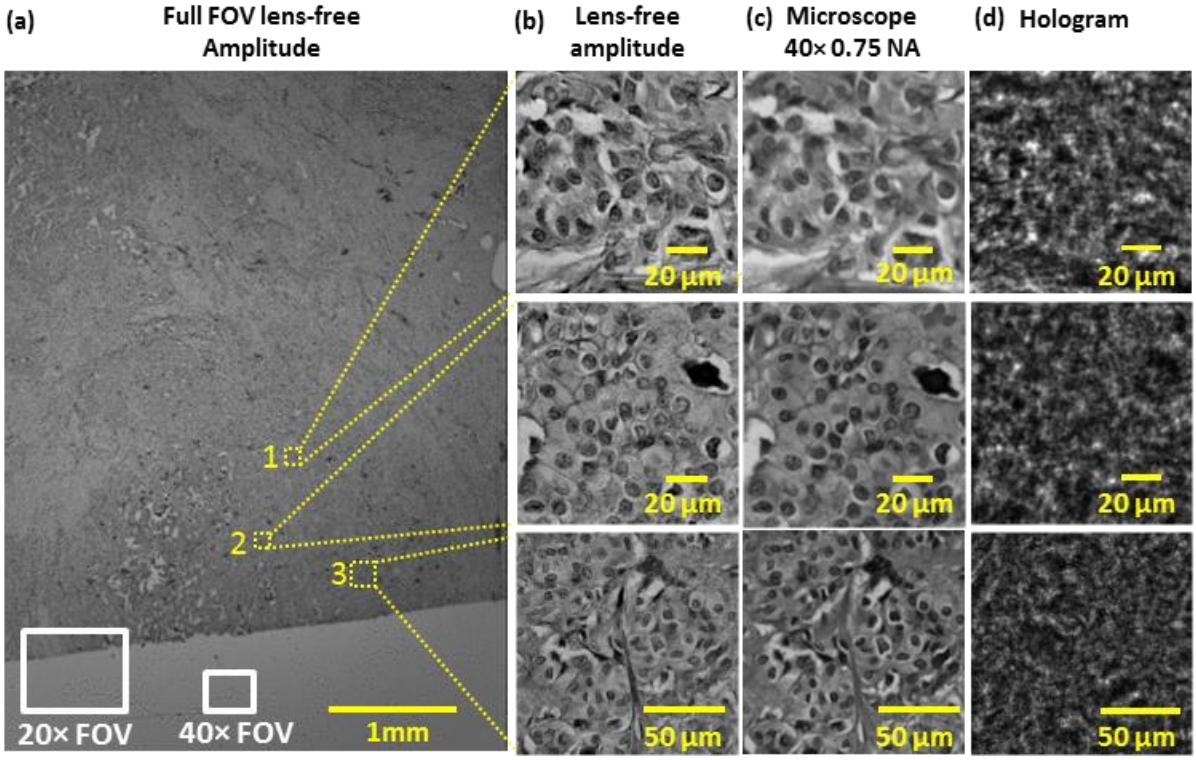

Figure 1.5 Lens-less imaging of human breast pathology slide. (a) Full-FOV lens-less amplitude image of the specimen. For comparison, the typical digital FOV of conventional 20× and 40× microscope objectives are shown by solid rectangles. (b) Zoomed-in regions outlined by the yellow squares in (a), showing a disordered epithelium. (c) Microscope comparison image of (b) taken with a 40× objective lens (0.75 NA). (d) Super-resolved lens-less holograms that are used to generate the images shown in (b). Reprinted from Ref. [34]

A recent work, inspired by the fact that almost all natural images, including those of biological samples, are sparse in the wavelet domain, has demonstrated that e.g., CDF9/7 wavelet encoding can be used to simplify the phase recovery process in lens-less on-chip microscopy [32]. This new approach, called sparsity-based multi-height phase retrieval, has demonstrated that high quality phase recovery and image reconstruction can be achieved using only two heights of hologram measurements for dense and connected biological samples, instead of the 6-8 heights employed in the regular multi-height phase retrieval method, giving a significant reduction in the number of measurements required (**Figure 1.6**). I will discuss other applications of sparse signal recovery and compressive sensing/sampling related approaches in computational lens-less digital holography in section 1.8.

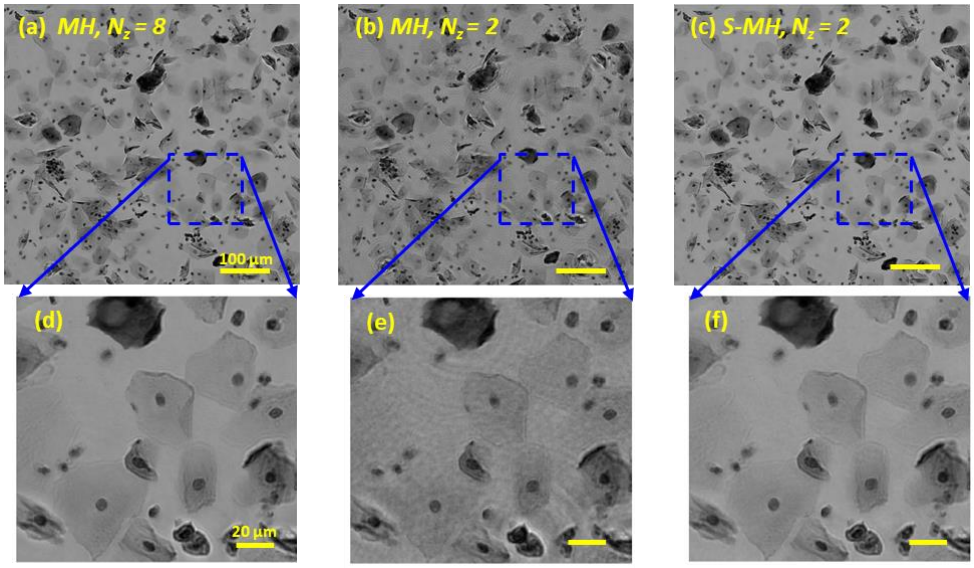

Figure 1.6 Sparsity based multi-height phase recovery (S-MH) achieves high quality phase recovery using 4× less hologram measurements than multi-height phase recovery (MH). (a) Reconstruction of the sample obtained by capturing holograms at $N_z = 8$ heights and processed by multi-height phase recovery method. This image also serves as the reference image for comparison. (b) Same as (a), except for $N_z = 2$. (c) Reconstruction obtained by capturing holograms at $N_z = 2$ heights and processed by sparsity-based multi-height phase recovery method, which shows similar quality as (a) while using 4-fold less measurements. (d-f) Zoomed-in regions of (a-c) marked by blue rectangles. Reprinted from Ref. [32]

As another degree of freedom, multiple angles of illumination can also be used for phase recovery, while at the same time increasing the effective NA of a reconstructed image through the synthetic aperture approach [38,54]. In addition to iterative algorithms as described above, there are also recursive phase retrieval methods that use Kalman filtering and multiple out-of-focus measurements [82,83]. Kalman filtering based recursive phase retrieval method offers good noise-reduction performance and can robustly reconstruct objects at low SNR. A later work has also simplified the computational cost of this Kalman filtering based approach using a diagonal approximation of the correlation matrix [84]. However, these methods require measurements (e.g., 50-100 heights) compared to iterative multi-height phase recovery techniques (e.g., two heights) and their performance for high resolution imaging of dense and connected biological samples is relatively inferior in lens-less on-chip imaging.

1.8 Compressive sensing/sampling and sparsity-based phase recovery

Compressive sensing (or compressive sampling) framework aims to reconstructs a signal (x) using measurements (y) of much smaller dimension ($\dim(x) \gg \dim(y)$), i.e., the measurement system, A, is under-determined. The basic assumption is that the signal to be reconstructed can be represented as a sparse function in some encoding domain. The compressive sensing/sampling inherently solves an optimization problem to recover a signal, x:

$$x^* = \arg\min \|Ax - y\|_p^q + \alpha \cdot \mathrm{card}(Cx) \tag{1.18}$$

where α is a regularization parameter, C is the encoding matrix, $\mathrm{card}(\cdot)$ is the cardinality function, which minimizes the number of non-zero elements in Cx.

Under some assumptions [52], the above non-convex problem can be equivalent to the following convex optimization problem:

$$x^* = \arg\min \|Ax - y\|_p^q + \alpha \cdot \|Cx\|_1 \tag{1.19}$$

The compressive sensing/sampling framework requires that the sought signal is sparse in some encoding domain $\mathrm{range}(\mathrm{C})$, and that the basis of the measurement matrix A is uncorrelated with that of the encoding matrix C. The form of the encoding matrix C depends on the *a priori* information on the sample or its image. For instance, if the image/sample itself is sparse (e.g., a fluorescently tagged object), then $\mathrm{C} = \mathrm{I}$ would be sufficient, where I is the identity matrix. As another example, if the image's derivative is sparse (i.e., only a few jumps in the image), then C can simply be selected as a 2D gradient operator D, and $\mathrm{TV}(\mathrm{x}) = \|\mathrm{Dx}\|_1$ is usually called the total variation or TV-norm [85]. In statistics and machine learning literature, this is also known as the least absolute shrinkage and selection operator (LASSO) method [86,87].

Using compressive sensing/sampling, digital holographic phase recovery performance can be significantly improved [88]. Some of the early work shows that the wave propagation itself is an efficient encoding scheme in compressive sensing, which enables e.g., single-height digital holographic phase recovery without the need for an object mask [89], and 3D sectioning of samples from 2D holographic measurements, among other achievements [90,91]. However, these schemes that use wave propagation as a sparse encoder mostly work for objects that are relatively isolated and sparse, and cannot work with dense and connected biological samples, which is an important application of lens-less microscopy. As discussed in the earlier sub-section, sparse signal recovery can also be used to image dense objects using a lens-less on-chip microscope by merging it with multi-height based iterative phase recovery to reduce the number of heights that is required to two heights. Compressive sensing has also been used in lens-less multi-spectral imaging, to demultiplex images obtained with a wavelength-multiplexed illumination [92], and to significantly increase the resolution of lens-less fluorescence on-chip microscopy and undo the effects of diffraction [50,93–95].

1.9 Color imaging in lens-less digital holographic microscopy

In many applications, including e.g., biomedical related ones, a color image is preferred, as it provides additional information and contrast of the sample. However, holography uses a quasi-monochromatic illumination for temporal coherence considerations. To get a color image of a sample, multiple holograms under different illumination wavelengths can be acquired and digitally combined to synthesize a color image. For instance, holograms using illumination wavelengths at the red (R), green (G) and blue (B) parts of the spectrum can be sequentially recorded, reconstructed separately and then digitally merged into a color image [20,38]. This requires 3-fold increase in data acquisition compared to that of a monochrome reconstruction.

One method to have less number of measurements for color imaging is called YUV averaging [40,96]. In this approach, high resolution and high-quality images (through PSR and phase recovery, i.e., without pixelation and twin image artifacts) are only acquired at a single illumination wavelength to create the Y channel (brightness) of a color image. Two additional low-resolution holograms are captured, back-propagated, and merged to get the U and V channels (color) of the image. Unfortunately, since this YUV averaging method smears the color information carried out in the U and V channels, color leakage is observed at the borders of an object (e.g., a stained cell). Moreover, because the Y channel comes from a single wavelength, the brightness of the image is biased by the transmission of the sample at that wavelength (see e.g., the third column of **Figure 1.7**).

Another method to improve the color imaging efficiency of an on-chip holographic microscope is to *simultaneously* use three illumination wavelengths, which effectively generates a multiplexed hologram that is formed by the summation of each hologram intensity at different color channels. This multiplexed hologram, captured through a Bayer color filter array (CFA) on a color image sensor chip, can be de-multiplexed through a matrix inverse operation using the

pre-calibrated spectrum of the sensor response [33,97–99]. However, because the pixels of different color channels on a Bayer CFA are not at the same location, conventional demoisaicing methods that rely on interpolation of different channels generate color artifacts at rapidly oscillating holographic fringes. This color artifact is much more pronounced in holography because a hologram contains high frequency fringes and through digital wave propagation, these localized color artifacts can permeate to the entire object image. This fringe-related color artifact can be digitally mitigated by using PSR to generate virtual pixels that super-impose upon each other. This approach is known as demosaiced pixel super-resolution (D-PSR) method, and has been shown to generate high-fidelity color images of stained biological samples using multi-wavelength illumination with sources that are *simultaneously* on [33,49]. As an example, **Figure 1.7** demonstrates the imaging of color-stained Pap smears using D-PSR, YUV averaging, and sequential RGB illumination in holographic imaging, where D-PSR achieves similar color imaging performance compared to RGB sequential but requires three folds less number of measured holograms. An alternative technique for color imaging in a holographic on-chip microscope combines a lens-less microscope with a mobile-phone based microscope, through wavelet fusion of mobile-phone microscope's low-resolution color images with a pixel super-resolved lens-less reconstruction at a single wavelength [100].

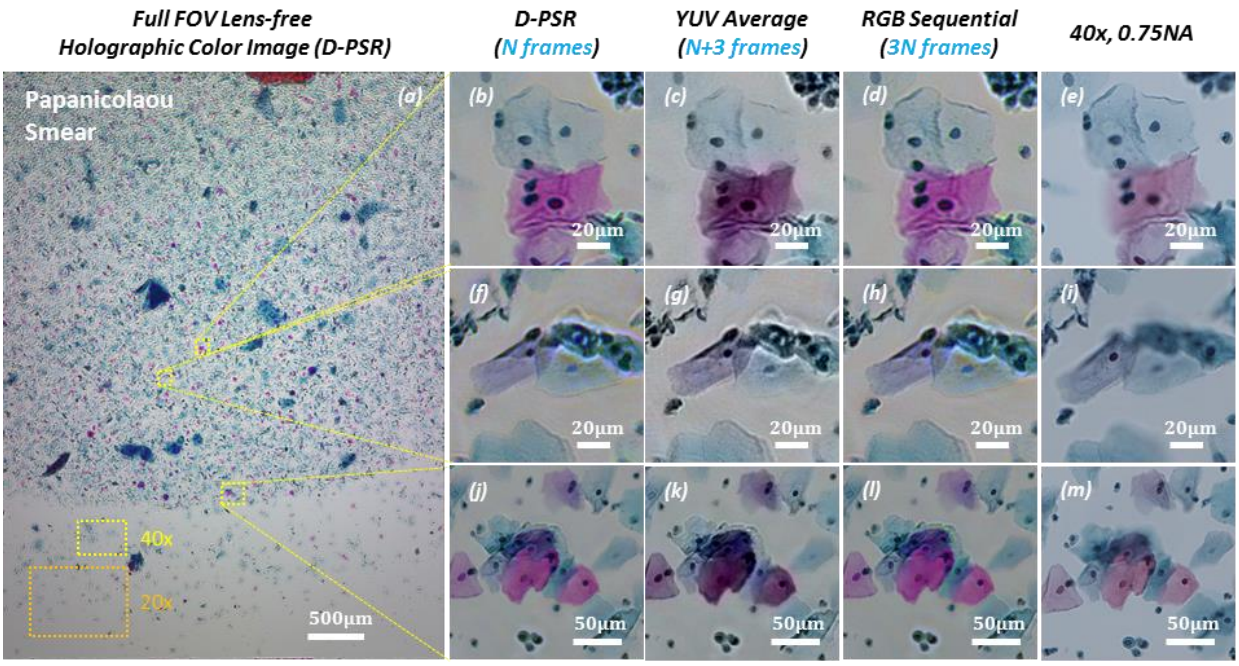

Figure 1.7 Wide field-of-view lens-less holographic color imaging of a Papanicolaou Smear. (a) Full field-of-view lens-less holographic image that is reconstructed using D-PSR under wavelength-multiplexed illumination at 470 nm, 527 nm and 624 nm. N = 144 raw holograms are used for this reconstruction. (b,f,j) Zoomed-in regions of (a). (c,g,k) Same zoomed in regions-of-interest reconstructed using the YUV color-space averaging method [96]; N+3 = 147 raw holograms are used. YUV color-space averaging method shows intensity bias and color leakage artifacts. (d,h,l) Same zoomed in regions-of-interest reconstructed using sequential RGB illumination [38]; 3N = 432 raw holograms are used. (e,i,m) Lens-based microscope images of the same samples are provided for comparison. These microscope images are blurred in some regions due to limited depth-of-focus of the objective-lens compared to lens-less holographic imaging. Typical FOVs of a 40x and a 20x objective-lens are also shown in (a) for comparison with lens-less microscope. Reprinted from Ref. [33]

Chapter 2 Deep learning in computational microscopy

Deep learning is a state-of-art machine learning technique to automatically analyze signals or data, which uses multi-layered neural networks. The word "deep" refers to the fact that there are tens, sometimes hundreds, of layers in the neural networks; in fact, it has been shown that a deeper network can perform better for many modern tasks [101,102,102,103]. The basic building block of these neural networks, which is also known as artificial "neurons", is a combination of a linear filter and a nonlinear activation function, which, when put together, can approximate an arbitrary function [104]. Neural network was first proposed in 1940s [105], and many fundamentals of neural network were established around 1990s [106–109]. Recently, there is a renaissance of deep neural networks, which is powered by large amount of data, advancements in parallel computation power and various open-source frameworks [110–113]. Deep neural networks have been proven to be very effective in a variety of tasks including image classification [101–103,107,114], natural language processing [115] and playing strategic games [116], among others [117,118].

More recently, deep learning has been used for many applications in computational microscopy, solving inverse problems for image restoration, which I will detail in section 2.1. Among other neural network architectures, convolutional neural network (CNN) have been shown to be very effective in processing images, partially thanks to their shift-invariant properties. In section 2.2, I will introduce the design and implementation of CNNs, which will be used throughout this dissertation. Based on these fundamentals, in sections 2.3, 2.4 and 2.5, I will detail my inventions and contributions in deep learning enabled three-dimensional computational microscopy, for applications in digital holography, bright field microscopy and fluorescence microscopy, respectively.

Part of this chapter have been previously published in:

- Y. Wu et al., “Extended Depth-of-Field in Holographic Imaging Using Deep-Learning-Based Autofocusing and Phase Recovery.” Optica 5, no. 6 (June 20, 2018): 704–10.;
- Y. Wu et al., “Bright-Field Holography: Cross-Modality Deep Learning Enables Snapshot 3D Imaging with Bright-Field Contrast Using a Single Hologram.” Light: Science & Applications 2019;
- Y. Wu et al., “Three-Dimensional Propagation and Time-Reversal of Fluorescence Images.” Nature Methods 2019 (accepted);
- Y. Rivenson, Y. Wu, and A. Ozcan. “Deep Learning in Holography and Coherent Imaging.” Light:Science&Application 8, no. 1 (September 11, 2019): 1–8.;
- K. Hann, Y. Rivenson, Y. Wu, and A. Ozcan., “Deep learning-based image reconstruction and enhancement in optical microscopy.” Proceeding of IEEE 2019 (accepted).

2.1 Image restoration and reconstruction using deep learning

Chapter 1 introduced the modeling of imaging system by PSF, as well as iterative algorithms such as error reduction and compressive sensing etc. to solve inverse problems (equation (1.2)) for various image restoration and reconstruction tasks including pixel super resolution and phase recovery. Here, I will introduce how deep learning can significantly improve the results of similar tasks, and enable novel tasks beyond the capability of conventional algorithms.

2.1.1 Single image super resolution

Optical microscopes have finite aperture, which limits the extent of spatial frequencies that can pass through the system, and therefore defines the smallest resolvable feature d:

$$\mathrm{d} = \frac{\lambda}{2\mathrm{NA}} \tag{2.1}$$

where λ is the wavelength of illumination, and NA is *numerical aperture* of the imaging system, which defines the ability of the lens to gather diffracted object light from a fixed distance, and cannot exceed the refractive index of the sample medium. For instance, a fundamental diffraction

limit is $\lambda/2$ in the air (refractive index 1), which restricts resolution in the visible range to be around 200-300 nm, unless other super resolution techniques that beats the diffraction of light are employed. Another resolution limit in optical microscopy comes from the finite spatial sampling of the image sensor chip, where the smallest resolvable feature cannot exceed twice the effective pixel size of the system, as has been detailed in section 1.4.

As described in section 1.1, these effects can all be attributed together to the system PSF, which in some cases can be assumed to be spatially invariant (equation (1.3)), but in general is spatially variant due to system aberrations introduced by various optical components as well as the mismatch of the sample refractive indices and the optical medium (equation (1.1)). The *inverse problem* of such PSF-related resolution degradation is known as super resolution by deconvolution, which aims to recover the missing spatial frequencies in a microscopy system by solving the inverse problem. Although this problem in general is ill-conditioned, because an infinite set of high-resolution images can all degrade to the same low resolution image (i.e. spatial frequencies beyond diffraction limit can be arbitrarily extrapolated), such extrapolation in theory can be feasible based on the measured (or a-priori known) spatial frequencies of an object[119], using, for example, the principle of analytical continuation [120]. In practice, the success of *frequency extrapolation* is fundamentally related to the imaging system's signal to noise ratio (SNR)[119] and a-prior knowledges; in fact, better results can be achieved with more accurate a-priori knowledge of the signal (e.g. sample image properties and structures) and system PSF.

A deep neural network acquires such a-priori knowledge through training with accurately registered image pairs of low-resolution images as input and corresponding high resolution images as ground truth label. After training, the deep neural network can perform a statistical inference acting on the low-resolution input images to match the corresponding high-resolution

labels. For microscopy applications, this form of super-resolution is far reaching as it allows for larger fields of view to be measured per acquired image, which can be particularly useful for high throughput imaging. In other words, low NA objective lenses can image a much larger field of view as illustrated in **Figure 2.1(a)**, thus the technique improves over the native space-bandwidth product of an objective lens [121]. This super-resolution image enhancement has been demonstrated for fluorescence, bright-field, and coherent (holographic) imaging systems, as illustrated in **Figure 2.1(b-d)**. In many cases, the deep networks were also able to generalize to different types of samples that were not part of the training set. For example, super-resolution of bright-field microscopy images that were trained with one tissue type and was successfully tested on another [122]. Recently, a unique framework termed as *Deep-Z* uses deep learning to *selectively* deconvolves the spatial features that come into focus, while convolving other features that go out-of-focus, endowing snapshot *3D* refocusing capability of coherent imaging and holography to incoherent fluorescence microscopy [123]. I will expand *Deep-Z* approach in section 2.5.

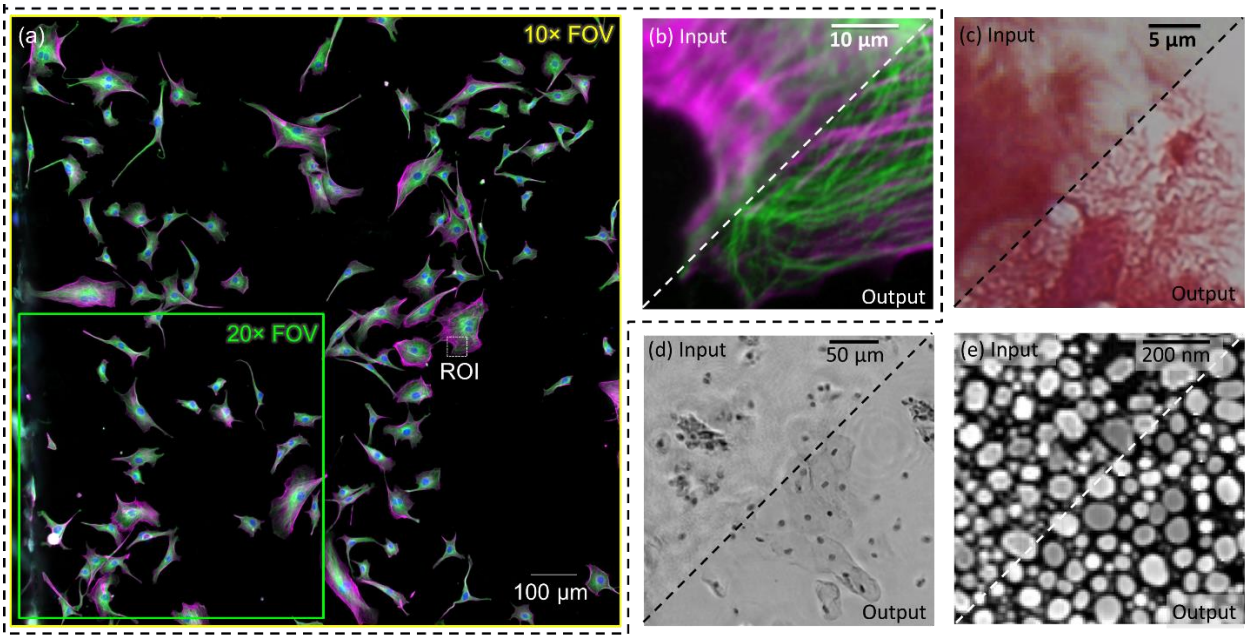

Figure 2.1 Super-resolution and image quality enhancement using deep learning. (a) A fluorescence microscopy image captured with a 20× objective lens and 10× field-of-view (FOV) marked. (b) Zoomed in region of

(a) showing a network perform super-resolution of a 10×/0.4NA input image. Images taken from Ref. [124]. (c) Super-resolution of a bright-field microscope image. Input image is taken using a 40×/0.95NA microscope objective. Images taken from Ref. [122]. (d) Super-resolution of an image created using digital holographic microscopy. Images taken from Ref. [125]. Input is imaged using a 4×/0.13NA microscope objective. Images taken from Ref. [125]. (e) Super-resolution of a scanning electron microscope (SEM) image. Images taken from Ref. [126].

In addition to light microscopy, similar techniques have also been applied to enhance images acquired by a scanning electron microscope (SEM) [126] (see **Figure 2.1(e)**). SEM imaging is not limited by the wavelength of electrons, but instead by aberrations and pixel size of the imaging system. For the latter, de-aliasing through deep learning can be used to perform pixel super-resolution of electron microscopy images, so that the electron doses and related sample damage can be reduced, without sacrificing adequate SNR. SEM can also be expensive and slow to image large areas at the nanoscale level, which could be improved using the same method.

2.1.2 Phase recovery and hologram reconstruction

As discussed in section 1.5, one of the most important and challenging tasks in digital holography is phase recovery, since the opto-electronic sensor is only sensitive to the intensity of the impinging light. Many optical and/or numerical methods have been proposed for phase recovery analytically or iteratively [34] (see e.g. sections 1.6, 1.7 and 1.8 for details). In contrast to these physics-driven hologram reconstruction approaches, emerging data-driven alternatives based on deep learning have recently demonstrated rapid and robust holographic image reconstruction from a single hologram. These data-driven approaches use accurately-registered and labeled image data to train a CNN; these high-quality image labels used as the ground truth for the training phase can be obtained from e.g., known sample structures [127] (**Figure 2.2(a)**), or by using a physics-based iterative reconstruction method [128] (**Figure 2.2(b)**). After its training,

the network can blindly transform a distorted, low-quality image obtained from a single hologram intensity into the desired high-quality label/image [127,128] (**Figure 2.2**). In general, a better reconstruction quality can be achieved through *physics-based learning* approaches, for example, by first refocusing the hologram (without phase recovery) onto the object plane, and then using deep learning based inference (see e.g., **Figure 2.2(b-d)**).

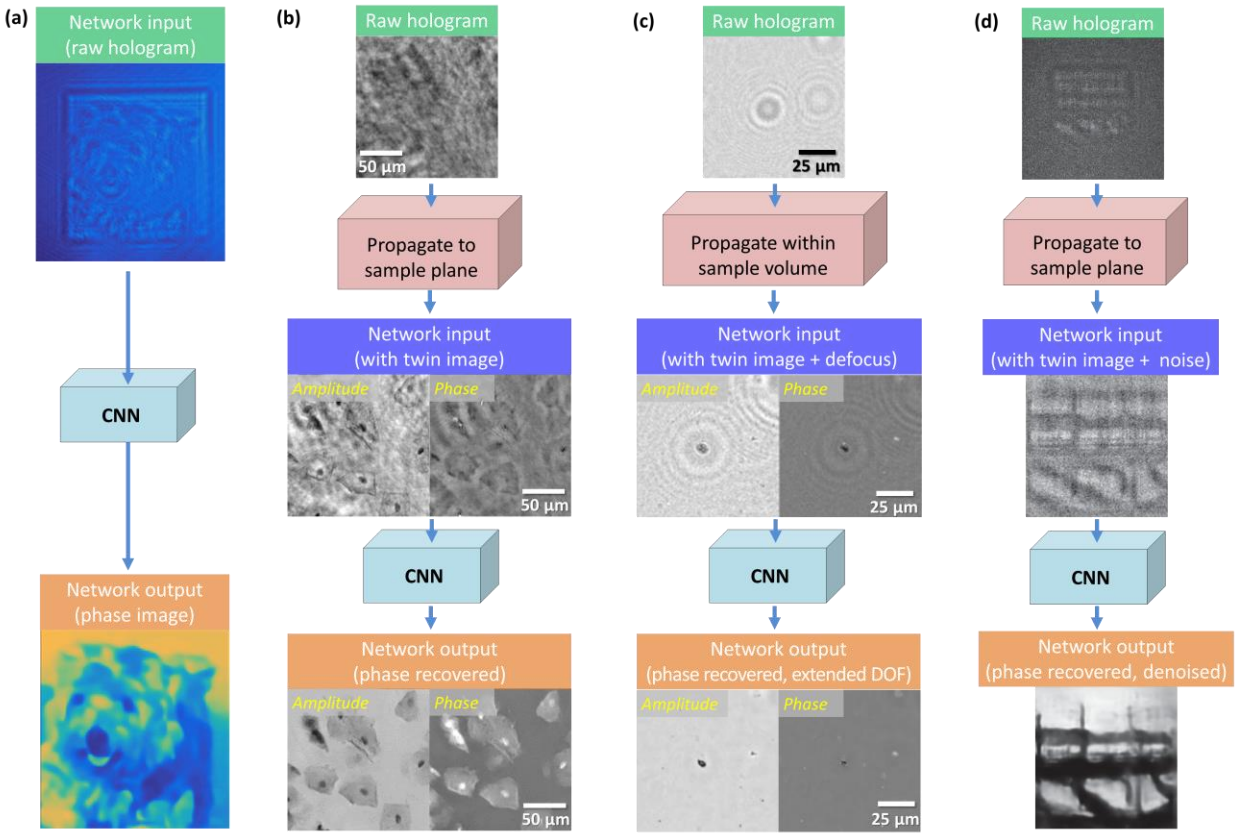

Figure 2.2 Deep learning-based hologram reconstruction. (a) An end-to-end CNN was trained to transform a hologram directly to a phase image[127]. (b) The raw hologram (i.e., without phase information) was numerically focused to the sample plane and was used as an input for the network to match the phase-recovered image[128]. The sample is a Pap smear specimen. (c) The raw hologram was propagated to an approximate distance within a sample volume, and the deep network generated an extended depth-of-field reconstruction of the whole sample volume, also performing auto-focusing[129]. The specimen is a 3D distributed aerosol sample. (d) Similar to (b) but implemented on holograms under low-photon and poor SNR conditions[130]. Adapted with permission from Ref. [131]

A recent work[129] further demonstrated the ability of a trained deep neural network to perform, *simultaneously*, auto-focusing and phase recovery to generate an extended depth-of-field image from a single hologram measurement (**Figure 2.2(c)**). In this deep learning-based framework, which is termed *Holographic Imaging using Deep Learning for Extended Focus* (HIDEF), the network is trained using pairs of randomly defocused (backpropagated) holograms and their corresponding in-focus, phase-recovered images. HIDEF significantly decreases the time complexity of holographic image reconstruction in 3D through simultaneous refocusing and phase recovery of 3D distributed sample points, which is done all in parallel. I will expand this HIDEF approach in section 2.3.

Deep learning has also been used to perform denoising of object images reconstructed from their holograms, to substantially increase the SNR of the output images[130,132] (**Figure 2.2(d)**). In one of these approaches [132], the network was trained using high SNR images as the labels, along with computationally simulated input images that had lower SNR. The trained network was then used to blindly perform robust speckle noise reduction in experimentally obtained image data. A similar framework [130] was also used to successfully retrieve the shape of an object from its photon-starved hologram with an SNR that is close to one.

In addition to being more data efficient (using only a single hologram measurement), these deep learning-based hologram reconstruction methods also demonstrate a four- to thirty-fold increase in the reconstruction speed compared with the state-of-the-art iterative hologram reconstruction algorithms [129,130].

2.1.3 Imaging through scattering media and diffraction tomography

The applications of deep learning in coherent imaging systems are not limited to holography, which assumes a *single-scattering event*. Using accurately-labeled and cross-registered datasets

of input-label image pairs, a deep neural network can also be trained to digitally reverse *multiple-scattering events* and reconstruct a sample's image even through scattering media, where the scattering event can also be approximated by a system PSF in the forward model defined as in section 1.1. For example, a deep neural network was trained for image reconstruction through glass diffusers under coherent illumination [133,134] (**Figure 2.3(a)**). A related method was also demonstrated to reconstruct and classify handwritten digits from input images of speckle patterns, obtained through multimode fiber propagation over a distance of up to 1 km[135] (**Figure 2.3(b)**).

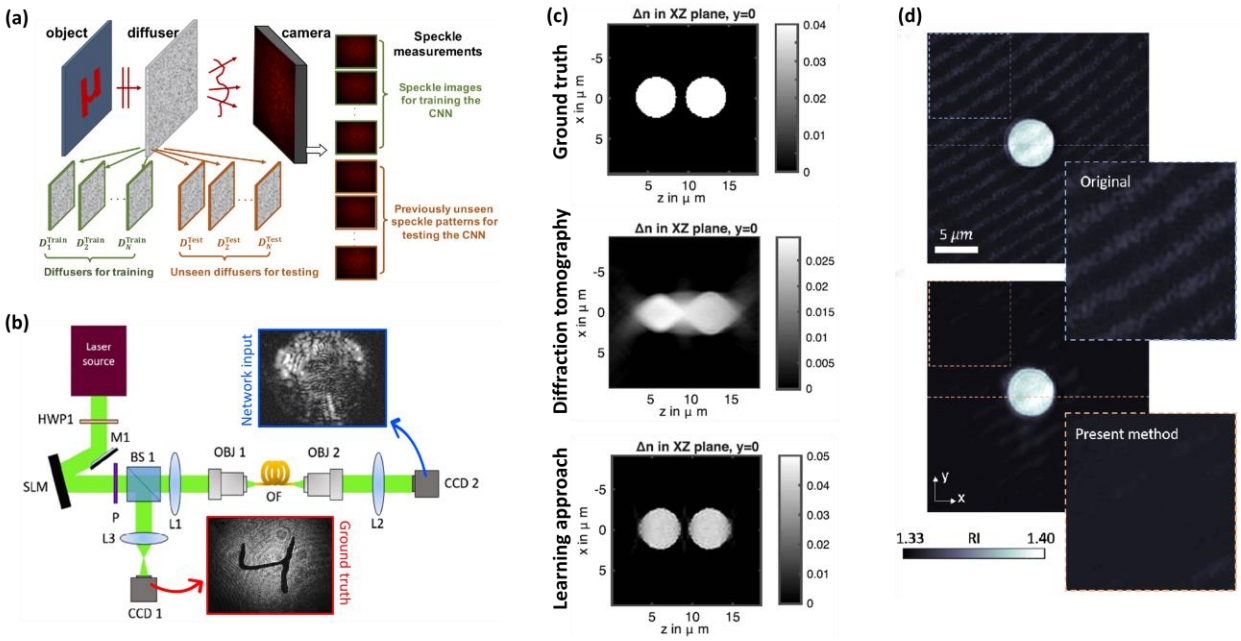

Figure 2.3 Deep learning-based imaging through complex media. (a) Imaging through a diffuser[134]. Reprinted with permission from Ref. [134]. (b) Imaging through a multi-mode fiber[135]. (c) Learning-based diffraction tomography reconstruction[136]. (d) Speckle noise reduction in diffraction tomography[137]. Reprinted from Ref. [131]

Deep learning has also been applied to optical diffraction tomography. In one of the earlier studies in this field, Kamilov *et al.* demonstrated that a trained, fully connected neural network can form an inverse model to reconstruct the 3D refractive index distribution of cells from diffraction tomography recordings [136] (**Figure 2.3(c)**). Recently, it has been also demonstrated

that a CNN can be trained and used for an ill-conditioned inverse imaging problem, providing tomographic reconstruction of densely layered objects from limited angle intensity projection recordings [138]. Another recent work has successfully utilized a generative adversarial network (GAN) for reducing dynamic speckle noise in diffraction tomography images (**Figure 2.3(d)**), using unregistered pairs of input and label images during the training process [137].

2.1.4 Transformation between microscopy modalities and contrast mechanisms

Another interesting avenue is to learn a statistical transformation between two *different* microscopic imaging modalities and/or contrast mechanisms, which allows the network to achieve results that are *not possible using standard forward model-based inverse problem solutions*. Using this deep learning-based approach, cost-effective or simpler microscopes can take the same quality of measurements as the gold-standard microscopes, helping to democratize microscopy related research and innovation. An example is lens-less digital holographic microscope, which has the advantage of compactness, cost-effectiveness, snapshot volumetric imaging and wider field-of-view. However, the image quality and contrast of a lens-less microscope is inferior compared to a benchtop bright field microscope due to various sources of coherent noise. Using cross-modality deep learning transformation, I developed a framework that transforms lens-less holography image to the equivalence of benchtop bright field image (**Figure 2.4(b)**) [139]. This unique framework is known as "bright field holography", which I will expand in section 2.4. Similar transformation has also been demonstrated for low-cost and portable cellphone microscopes, which corrects the aberrations and improves the resolution and SNR of a cellphone microscope image (**Figure 2.4(a)**) [140]. As another example, cross-modality image transformations can be used to enhance the resolution of a microscopy modality beyond the

diffraction limit, transforming diffraction limited confocal microscopy images into stimulated emission depletion (STED) microscopy equivalent images (**Figure 2.4(c)**)[141].

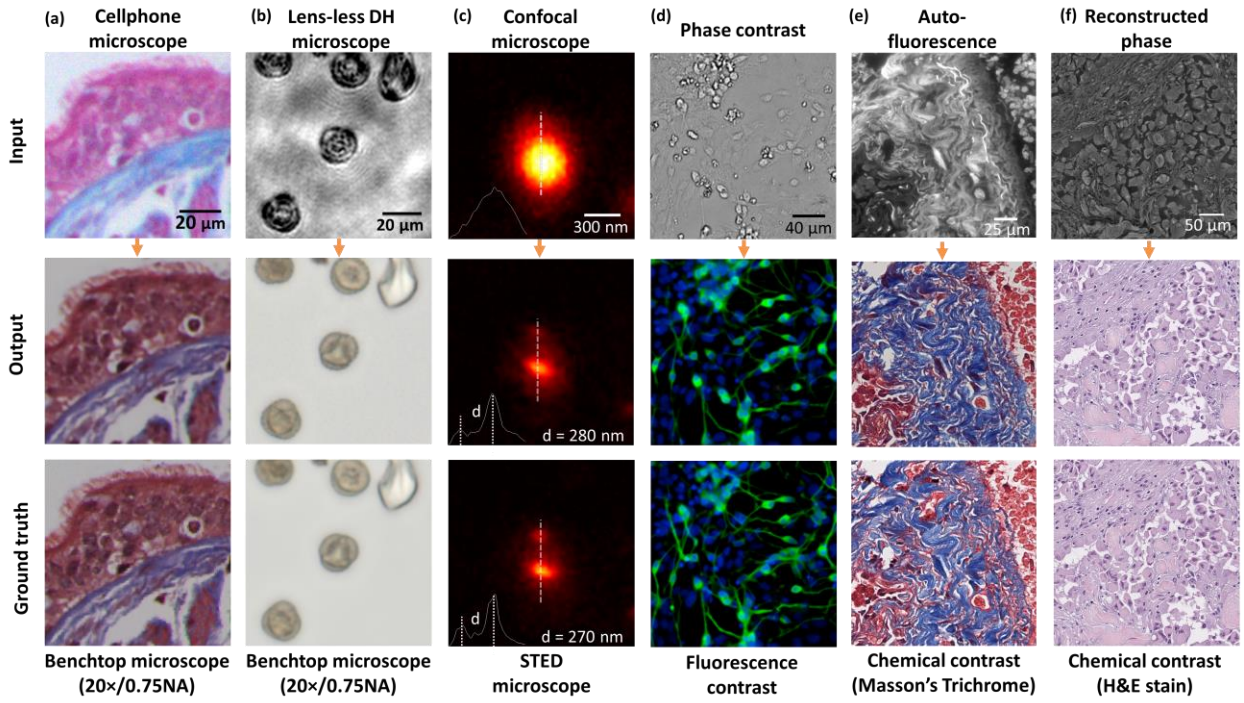

Figure 2.4 Transforming the images between different microscope modalities and contrast mechanisms. (a) Cellphone microscope image transformed into a benchtop microscope equivalent image. (b) Lens-less digital holographic (DH) microscope image transformed into a benchtop bright-field microscope equivalent image. (c) Confocal microscope image transformed into a stimulated emission depletion (STED) equivalent image. (d) Label-free phase contrast image stack transformed into fluorescence-labeled image. (e) Label-free auto-fluorescence image transformed into histologically stained bright-field microscope equivalent image. (f) Label-free quantitative reconstructed phase image transformed into histologically stained bright-field microscope equivalent image.

Using similar framework, deep learning has also been used to perform image transformations across different *contrast mechanisms*. One of the most interesting direction for this application is bypassing chemical labeling techniques (normally used to add exogenous contrast to biological samples), where deep learning can be used to transform one or more label-free images of the sample with indigenous contrast such as quantitative phase or auto-fluoresce into an equivalent

image of the chemically labeled sample. Such transformation has multiple benefits including: (1) saves the cost and labor for staining; (2) expedite the imaging and diagnostic process; (3) provides the possibility to virtually staining the *same* sample (such as a tissue section) by more than one contrast mechanisms; and (4) improve upon the repeatability of physical stains against lab-to-lab and technician-to-technician variations [142]. As an example, Rivenson *et al.*[143] demonstrated the efficacy of the deep learning-based staining using images of a single autofluorescence channel of an unstained tissue, and that the technique can be applied to several different tissue types and histological stains (**Figure 2.4(e)**). It was also demonstrated that additional excitation channels or imaging modalities can also be used to increase the staining quality, in cases where some of the tissue constituents do not provide meaningful contrast at a single band [143,144]. These techniques can also be used with other contrast mechanisms and microscope modalities, such as optical coherence tomography (OCT)[145,146], quantitative phase imaging (**Figure 2.4(f)**) [147], Raman microscopy [148,149] as well as other rapid staining methods [150–152]. As another example, deep learning was successful in predicting fluorescently-tagged neuron images using unlabeled quantitative phase images (phase contrast image stack) [153], enabling the detection of specific markers without the need of potentially lengthy, costly and toxic process of labeling (**Figure 2.4(d)**).

2.2 Network and training

The image restorations and transformations described above, as well as many others [154,155], are generally enabled by supervised optimization of CNNs using accurately-registered image pairs (**Figure 2.5(a)**). CNNs typically contain tens to hundreds of layers of convolution kernels (filters), bias terms, and non-linear activation functions. Through a process of *training* (which is a one-time effort), the weights of these filters and biases of the neural network are adjusted in a

way that minimizes the error between the network output image and the "gold standard" target labels, in terms of a user-defined cost or loss function (for example, mean squared error loss[106] **Figure 2.5(b)**). This trained network can subsequently be used to perform a pre-defined image reconstruction task with a single forward-pass through the network, yielding its inference. Such inference typically only takes a fraction of a second (using e.g., a standard graphics processing unit, GPU) without the need for manual tuning of any hyperparameters or a refinement of the physical assumptions made regarding the image reconstruction model; in fact, this non-iterative single forward-pass reconstruction capability forms one of the major advantages of deep learning-based solutions to inverse problems in imaging. In this section, I will introduce the strategies to build and implement a deep neural network for microscopic image reconstruction.

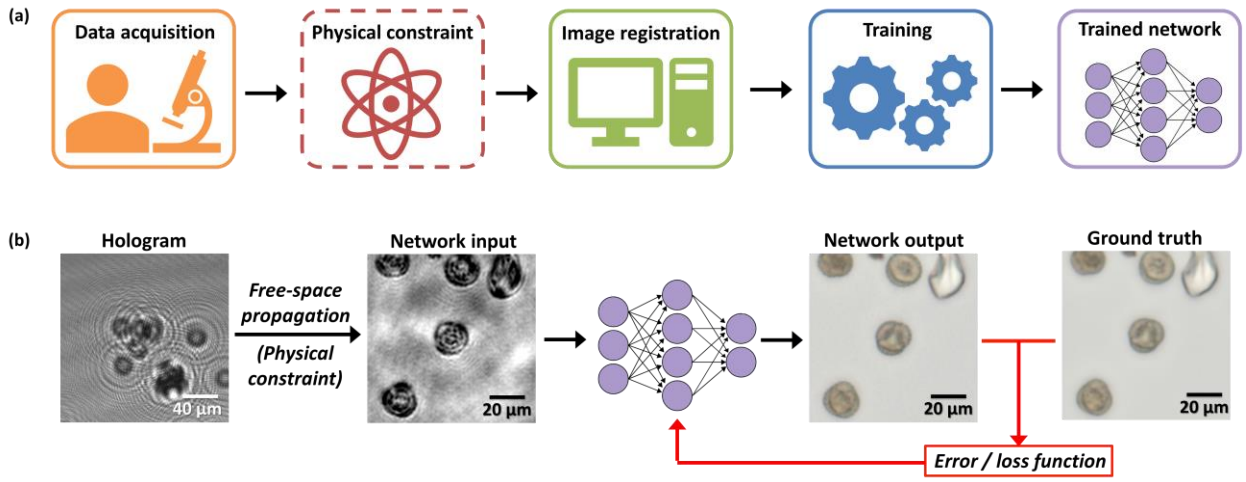

Figure 2.5 Training a deep neural network for image reconstruction. (a) Training workflow of an image reconstruction deep neural network, including data acquisition, physics-based constraints, image registration, and training. (b) An example network training and testing procedure, where the network learns to match the input image to a target label (ground truth image) using a given loss/cost function. Reprinted from Ref. [131]

2.2.1 Neurons, layers and network architecture

Several different types of layers are used by neural networks, including *fully-connected layers*, which connect between each neuron in two consecutive layers; *pooling layers*, which find

e.g., the maximum or average of adjacent tensor values; and *convolutional layers*, which provide a spatially limited connectivity across successive layers. Using all convolutional layers in a network, as in a CNN, has the property of parameter sharing and shift-invariance that enables homogenous detection of features within the image regardless of their locations, which is especially useful for microscopic image processing. It also allows for scaling of the network, where the testing can be performed on images of different sizes from training. Another desirable property of deep CNNs is that the convolutional layers can create hierarchical data representations [156]. Each convolutional layer maps successive data tensor x_i by convolution:

$$x_{i+1} = \mathrm{CONV}(x_i) = W_{i,j} \star x_i + b_{i,j} \tag{2.2}$$

where CONV stands for the convolution operator (including the bias terms), and $W_{i,j}$ and $b_{i,j}$ are the trainable convolutional filter and bias terms of layer i and channel j, respectively. The size of convolutional filter (e.g. 3×3) defines the receptive field of the convolutional layer, i.e. the size of the image feature that can be perceived by this layer. Larger receptive field can be achieved by cascading convolutional layers and/or pooling layers. In fact, it has been shown that cascading smaller filters (3×3) is more effective than using a larger filter (e.g. 5×5, or 7×7) directly [102].

A nonlinear activation function is typically used right after a convolutional or fully connected layer. While a number of nonlinear functions exist, *rectified linear unit (ReLU)* have become among the most popular [157], which is defined as:

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

ReLU is favorable because its gradient is cheap to compute and it mitigate the vanishing gradient problem [108], which facilitates highly efficient training of very deep neural networks [157–159].

The arrangement of layers into a neural network is called network *architecture*. A variety of network architectures have been developed; a few general types of them are most popular and

have been repeatedly demonstrated effective[103,160,161]. Regardless of architecture, some hyper-parameters of the network need to be tuned. An example is network size, which can be changed through the number of layers as well as number of channels in each layer of the network. Tuning these parameters can ensure that the network size is large enough to extract and learn important features but not so large that it overfits to training data [162,163]. For CNN, the size of convolutional and pooling layers also controls the effective receptive field and should be designed to sufficiently contain all the required information[164]. Hyper-parameters are often tuned manually by trying (cross-validating) different designs; an active area of research now is aiming automatic and optimal tuning of these hyperparameters of a neural network [165–167].

2.2.2 U-Net

One of the most popular network architectures for microscopic image transformations is *U-net* [161,168], which was originally used for image semantic segmentation but has been widely applied to image transformation ever since. **Figure 2.6(a)** illustrates an example U-Net, which has U shape and consists of a series of convolutional blocks and skip connections to allow the network to go deeper without vanishing gradients issues. It has a down-sampling path and a symmetric up-sampling path. In this example, there are five down-sampling blocks. Each block contains two convolutional layers that map the input tensor x_k to the output tensor x_{k+1}:

$$x_{k+1} = x_k + \mathrm{ReLU}\left[\mathrm{CONV}_{k_2}\left\{\mathrm{ReLU}\left[\mathrm{CONV}_{k_1}\{x_k\}\right]\right\}\right] \tag{2.4}$$

where the subscript of CONV denotes the number of channels in the convolutional layer; e.g., $k_1 = 25, 72, 144, 288, 576$ and $k_2 = 48, 96, 192, 384, 768$ for levels $k = 1, 2, 3, 4, 5$. The "+" sign in equation (1.10) represents a residual connection [103], which improves the convergence and performance of the CNN. The connection between two consecutive down-sampling blocks is a max-pooling layer (e.g., 2×2 pooling to perform a 2× down-sampling).

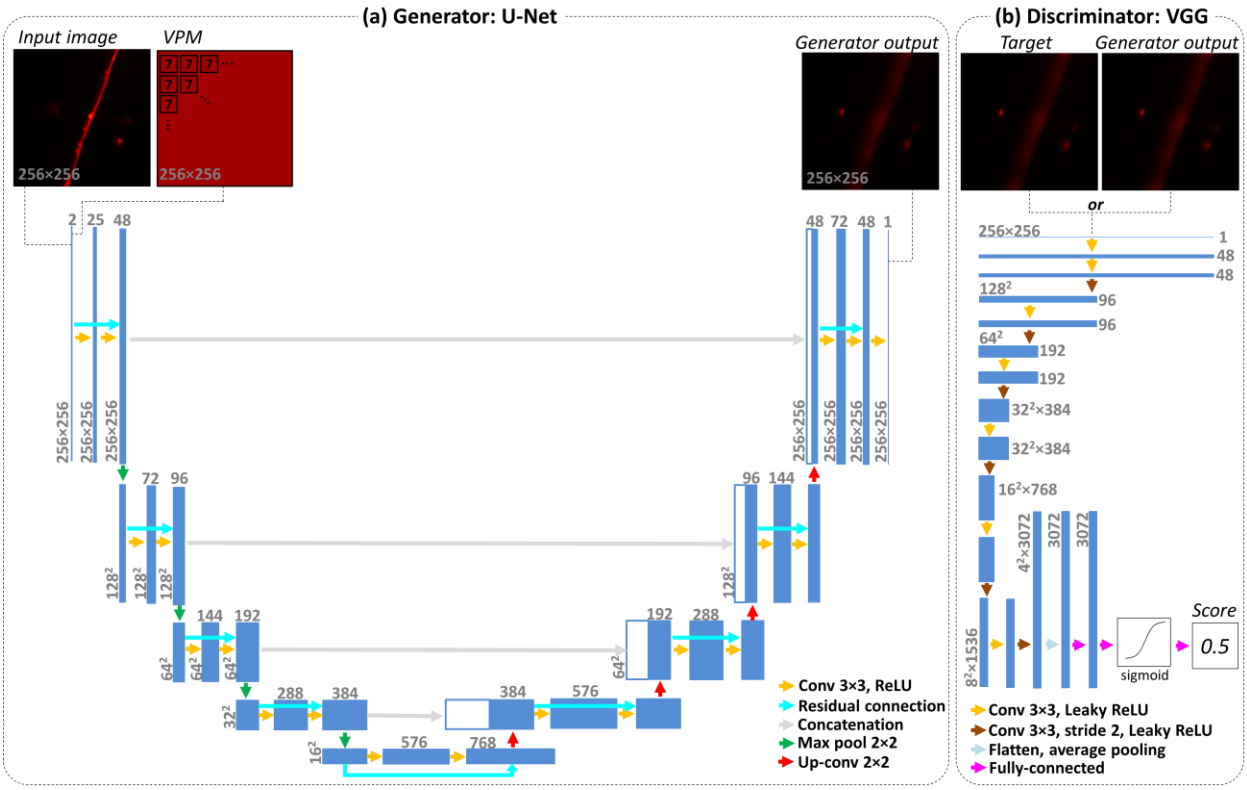

Figure 2.6 Generative adversarial network. (a) Generator network, using U-Net architecture. The network can also be trained and tested solely without discriminator as a CNN for image transformations. (b) Discriminator network, using VGG architecture. The network can also be trained and tested solely without generator as a CNN for e.g., image classifications. The numbers represent the size and the channels of each block. ReLU: rectified linear unit. Conv: convolutional layer. This network was used for *Deep-Z* (section 2.5).

The fifth down-sampling block connects to up-sampling path, in which there are four up-sampling blocks that are symmetric to the down-sampling path, each contains two convolutional layers that map the input tensor y_{k+1} to the output tensor y_k:

$$y_k = \mathrm{ReLU}\left[\mathrm{CONV}_{k_4}\left\{\mathrm{ReLU}\left[\mathrm{CONV}_{k_3}\{\mathrm{CAT}(x_{k+1}, y_{k+1})\}\right]\right\}\right] \quad (2.5)$$

where the $\mathrm{CAT}(\cdot)$ operator represents the concatenation of the tensors along the channel direction, k_3 and k_4 are the number of channels in the convolutional layers, e.g., $k_3 = 72, 144, 288, 576$ and $k_4 = 48, 96, 192, 384$ for levels $k = 1, 2, 3, 4$, respectively. This symmetric pairs of down-

sampling and up-sampling path potentially enables U-Net to separate and learn features at different spatial/size levels, whereas down-sampling allows a large receptive field. Different from a simpler encoder-decoder architecture[169], the additional skipping connections (marked in gray in **Figure 2.6(a)**) allows data to pass from down sampling to up sampling path at each level. U-Net, together with GAN (detailed below), forms the basic tool for the emerging computational microscopy approaches that I will expand in sections 2.3, 2.4 and 2.5.

2.2.3 Generative adversarial network (GAN)

Generative adversarial networks use two competing networks: a generator network $G(.)$and a discriminator network $D(.)$. Whereas U-Net can be used for generator for image transformation, an additional discriminator network can be jointly trained that discriminate between generated images G(x) and the ground truth label images z, by minimizing the discriminator loss function:

$$L_D = \frac{1}{2N} \cdot \sum_{i=1}^{N} \left[D\left(G\left(x^{(i)}\right)\right)\right]^2 + \frac{1}{2N} \cdot \sum_{i=1}^{N} \left[D\left(z^{(i)}\right) - 1\right]^2 \quad (2.6)$$

where *N* is the number of images used in each batch (e.g., N = 20), $G\left(x^{(i)}\right)$ is the generator output for the input $x^{(i)}$, $z^{(i)}$ is the corresponding target label, $D(.)$ is the discriminator. This loss function is used to teach the discriminator to distinguish between the generated images and the ground truth label images. On the other side, the generator learns to "fool" the discriminator to classify its outputs as ground truth, by minimizing the generator loss function:

$$L_G = \frac{1}{2N} \cdot \sum_{i=1}^{N} \left[D\left(G\left(x^{(i)}\right)\right) - 1\right]^2 \quad (2.7)$$

For image transformation tasks, additional terms such as mean absolution error (MAE) loss (see section 2.2.5 for details) can be added to the generator loss, to "condition" the network output such that the pixel location and structural information of input image is preserved [170]:

$$L_G = \frac{1}{2N} \cdot \sum_{i=1}^{N} \left[D\left(G\left(x^{(i)} \right) \right) - 1 \right]^2 + \alpha \cdot \frac{1}{2N} \cdot \sum_{i=1}^{N} MAE\left(x^{(i)}, z^{(i)} \right) \quad (2.8)$$

where α is a hyper-parameter that balances the generator and MAE loss (e.g. α =0.02). During training, L_G and L_D are iteratively updated, and the generator and discriminator competes, where both of their performance are improved. Adding a discriminator, with carefully tuned hyper-parameters, usually improves the image transformation result of the generator, which generates images with higher quality and resolution compared to CNN (i.e. training the generator only) [139].

The architecture of discriminator can be a classification CNN. For instance, the discriminator in **Figure 2.6(b)** is a VGG[102] style CNN that consists of 6 consecutive convolutional blocks, each of which maps the input tensor z_i to the output tensor z_{i+1}, for a given level i:

$$z_{i+1} = LReLU\left[CONV_{i_2}\left\{ LReLU\left[CONV_{i_1}\{z_i\} \right] \right\} \right] \quad (2.9)$$

where the LReLU stands for leaky ReLU operator with a slope of 0.01. The subscript of the convolutional operator represents its number of channels, e.g., $i_1 = 48,96,192,384,768,1536$ and $i_2 = 96,192,384,768,1536,3072$, for the convolution block $i = 1,2,3,4,5,6$, respectively. Subsequent flattening layers and fully connected layers were used to map the convolutional features into a discriminator score, which falls within (0, 1), using a Sigmoid activation function.

2.2.4 Data acquisition and image registration

Since supervised learning requires gold standard labels, it is important to acquire and register high-quality image datasets to train the deep network. CNNs have proven to generalize very well, but they are much more effective when trained on a dataset that is designed specifically for the test dataset that it will be used for. The training image datasets can be created in different ways. Most commonly, the input and gold-standard label images are imaged separately (by e.g., different optical microscope hardware) and then accurately registered to each other; alternatively,

the input images can be digitally generated using numerical degradation from label images (equation (1.2)), following e.g., a physical model or calibration of system PSF. These simulated datasets are typically much easier to create, but the efficacy of network performance depends on the accuracy of the forward model, which sometimes is a challenge to obtain beyond a simplified approximation. On the other hand, if both input and ground truth label images are experimentally obtained, they often need to be accurately registered to each other. Image registration for training data preparation is a one-time effort, which will be detailed below. A third method that can be used to generate training image data is to computationally reconstruct gold-standard label images. For instance, multi-height phase recovery (section 1.7) can be used to reconstruct phase recovered images using 4-8 hologram measurements, and a deep neural network can be trained to learn the transformation between a back-propagated single hologram to the computationally phase recovered image. This is used in HIDEF approach, as will be detailed in section 2.3.

When both the gold standard labels and the input images are experimentally acquired, different registration techniques can be used to spatially match these images to each other: (1) intensity-based registration (e.g. intensity cross-correlation), (2) feature-based registration (e.g. SIFT features) [171], and (3) deep learning based registration [172,173]. Whereas intensity-based registration typically works well even at the sub-pixel level when the two sets of images comes from similar imaging modalities, feature based registration methods can be more general across different modalities. In practice, a combination of registration steps may be needed. Since there can be huge variations between the images, especially when captured using different microscopy modalities, a unique workflow should be tailored for a given application of interest [174,147,140]. Section 2.5 explains a working example to accurately register two stack of images across wide field and confocal microscopies, enabling deep-learning based virtual refocusing across different

fluorescence microscopy modalities [123]. Note that some image transformation networks such as CycleGANs[175] do not need registered input and label images. However, without accurate registration for input-label image pairs, there is a possibility for artefacts.

Depending on the sample variation, the required training data size can range from a few thousand image patches to hundreds of thousands of patches or more. While deep learning has been shown to be highly effective on small datasets [176], a CNN in general works better when the training dataset is large enough such that the network can learn the entire sample space from it. However, for certain microscopy applications where either samples or images are rare and/or challenging to obtain, limited image data can lead to the network to overfit to the training data and not generalize well to test datasets. One popular method for improving the performance of neural networks with limited training data is *transfer learning* [177–179]. This process typically involves training a network on a large dataset that has similar features to the dataset that the network will be used on. Then using the trained parameters as parameter, the entire network or the last few layers of the network can be further tuned using the smaller dataset of interest. Another method is to augment the image dataset using techniques such as image rotation, flipping, shifting, and distortion [180], which also prevents the network from over-fitting.

Normalization of both the input image data and the ground truth labels can also significantly improve the consistency of the network inference. Most deep networks are highly *nonlinear*, meaning that small differences at the input images can cause relatively large differences at the network output. Therefore, any variations in the illumination source intensity or the exposure time of the microscopy system should be normalized during the training and testing phases.

2.2.5 Image quality evaluation

Several image quality evaluation criteria can be used as loss functions *during* supervised training, or for evaluating the output of the network *after* training. For registered image pairs, the

network output images I^{out} can be were evaluated with reference to the corresponding ground truth images I^{GT} using e.g., mean square error (MSE), root mean square error (RMSE), MAE, correlation coefficient, and structural similarity (SSIM) [181]. MSE is one of the most widely used error metrics, defined as:

$$\mathrm{MSE}(I^{out}, I^{GT}) = \frac{1}{N_x \cdot N_y} \left\| I^{out} - I^{GT} \right\|_2^2 \tag{2.10}$$

where N_x and N_y represent the number of pixels in the x and y directions, respectively. The square root of MSE results in RMSE. Compared to MSE, MAE uses 1-norm difference (absolute difference) instead of 2-norm difference, which is less sensitive to significant outlier pixels:

$$\mathrm{MAE}(I^{out}, I^{GT}) = \frac{1}{N_x \cdot N_y} \left\| I^{out} - I^{GT} \right\|_1 \tag{2.11}$$

The correlation coefficient is defined as:

$$\mathrm{corr}(I^{out}, I^{GT}) = \frac{\sum_x \sum_y \left(I_{xy}^{out} - \mu_{out}\right)\left(I_{xy}^{GT} - \mu_{GT}\right)}{\sqrt{\left(\sum_x \sum_y \left(I_{xy}^{out} - \mu_{out}\right)^2\right)\left(\sum_x \sum_y \left(I_{xy}^{GT} - \mu_{GT}\right)^2\right)}} \tag{2.12}$$

where μ_{out} *and* μ_{GT} are the mean values of the images I^{out} *and* I^{GT} respectively.

While these criteria can be used to quantify errors in the network output compared to the GT, they are not strong indicators of the *perceived similarity* between two images. SSIM aims to address this shortcoming by evaluating the structural similarity in the images, defined as:

$$\mathrm{SSIM}(I^{out}, I^{GT}) = \frac{(2\mu_{out}\mu_{GT} + C_1)(2\sigma_{out,GT} + C_2)}{(\mu_{out}^2 + \mu_{GT}^2 + C_1)(\sigma_{out}^2 + \sigma_{GT}^2 + C_2)} \tag{2.13}$$

where σ_{out} *and* σ_{GT} are the standard deviations of I^{out} *and* I^{GT} respectively, and $\sigma_{out,GT}$ is the cross-variance between the two images.

In addition to these measures, images quality can also be evaluated using a non-referenced measure, such as Blind/Reference-less Image Spatial Quality Evaluator (BRISQUE) [182].

2.3 Extended depth-of-field holographic imaging using deep learning

Holography encodes the three-dimensional (3D) information of a sample in the form of an intensity-only recording. However, as described in Chapter 1, to decode the sample image from its hologram(s), auto-focusing and phase-recovery are needed, which are in general cumbersome and time-consuming to digitally perform. Here I develop a convolutional neural network (CNN) based approach that simultaneously performs auto-focusing and phase-recovery to significantly extend the depth-of-field (DOF) and the reconstruction speed in holographic imaging. For this, a CNN is trained by using pairs of randomly de-focused back-propagated holograms and their corresponding in-focus phase-recovered images. After this training phase, the CNN takes a single back-propagated hologram of a 3D sample as input to rapidly achieve phase-recovery and reconstruct an in-focus image of the sample over a significantly extended DOF. This deep learning based method fundamentally improves the algorithm time-complexity of holographic image reconstruction from O(nm) to O(1), where n refers to the number of individual object points or particles within the sample volume, and m represents the focusing search space within which each object point or particle needs to be individually focused. The presented approach may also be used to computationally extend the DOF of other imaging modalities.

2.3.1 Introduction

Holography [2,16,17,34,35,183–185] encodes the three-dimensional (3D) information of a sample through interference of the object's scattered light with a reference wave. Through this interference process, the intensity of a hologram that is recorded by e.g., an image sensor, contains both the amplitude and phase information of the sample. Recovery of this object information over a 3D sample space has been the subject of numerous holographic imaging techniques [27,41,183–187]. In a holographic image reconstruction process, there are two major steps.

One is *phase-recovery*, which is required since only the intensity information of the holographic pattern is recorded at a given digital hologram. In general, for an off-axis holographic imaging system [184–187], this phase-recovery step can be achieved relatively easier compared to an in-line holography set-up, at the cost of a reduction in the space-bandwidth product of the imaging system. For in-line holography, on the other hand, iterative phase-recovery approaches that utilize measurement diversity and/or prior information regarding the sample have been developed [4,17,32–34,38,71,75,76,188–190]. Regardless of the specific holographic set-up that is employed, phase-recovery needs to be performed to eliminate the twin-image and self-interference related spatial artifacts in the reconstructed phase and amplitude images of the sample.

The other crucial step in holographic image reconstruction is *auto-focusing*, where the sample-to-sensor distances (i.e., relative heights) of different parts of the 3D object need to be numerically estimated. Auto-focusing accuracy is vital to the quality of the reconstructed holographic image such that the phase-recovered optical field can be back-propagated to the correct object locations in 3D. Conventionally, to perform auto-focusing, the hologram is digitally propagated to a set of axial distances, where a focusing criterion is evaluated at each resulting complex-valued image. This step is ideally performed after the phase-recovery step, but can also be applied before it, which might reduce the focusing accuracy [63]. Various auto-focusing criteria have been successfully used in holographic imaging, including e.g., the Tamura coefficient[62], the Gini Index[64] and others[63,65–70]. Regardless of the specific focusing criterion that is used, and even with smart search strategies[23], the auto-focusing step requires numerical back-propagation of optical fields and evaluation of a criterion at typically >10-20 axial distances, which is time-consuming for even a small field-of-view (FOV). Furthermore, if the sample has multiple objects at different depths, this procedure needs to repeat for every object in the FOV.

Some recent work has also utilized deep learning to achieve auto-focusing. Z. Ren *et al.* formulated auto-focusing as a classification problem and used a convolutional neural network (CNN) to provide rough estimates of the focusing distance with each classification class (i.e., bin) having an axial range of ~3 mm, which is more appropriate for imaging systems that do not need precise knowledge of the axial distance of each object[191]. As another example, T. Shimobaba *et al.* used a CNN regression model to achieve continuous auto-focusing, also with a relatively coarse focusing accuracy of >5 mm[192]. In parallel to these recent results, CNN-based phase-recovery methods that use a single intensity-only hologram to reconstruct a two-dimensional object's image have also been demonstrated[127,127,128,128]. However, in these former approaches the neural networks were trained with *in-focus* images, where the sample-to-sensor (hologram) distances were precisely known *a priori* based on the imaging set-up or were separately determined based on an auto-focusing criterion. As a result, the reconstruction quality degraded rapidly outside the system depth-of-field (DOF); for example, for high resolution imaging of a pathology slide (tissue section), ~4 μm deviation from the correct focus distance resulted in loss of resolution and distorted the sub-cellular structural details.[128]

Here, I demonstrate a deep learning based holographic image reconstruction method that performs both auto-focusing and phase-recovery at the same time using a single hologram intensity, which significantly extends the DOF of the reconstructed image compared to previous approaches, while also improving the algorithm time-complexity of holographic image reconstruction from *O(nm)* to *O(1)*. This approach is termed as *HIDEF* (Holographic Imaging using Deep learning for Extended Focus) and it relies on training a CNN with not only in-focus image patches, but also with randomly de-focused holographic images along with their corresponding in-focus and phase-recovered images, used as reference. Overall, HIDEF

significantly boosts the computational efficiency and the reconstruction speed of high-resolution holographic imaging by simultaneously performing auto-focusing and phase-recovery and increases the robustness of the image reconstruction process to potential misalignments in the optical set-up by extending the DOF of the reconstructed images (**Figure 2.7**).

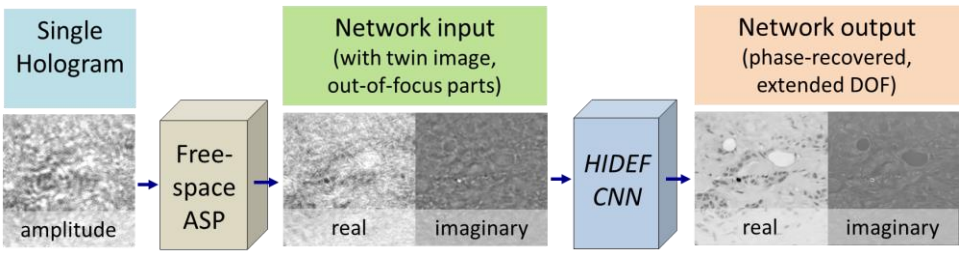

Figure 2.7 HIDEF achieves simultaneous phase-recovery and auto-focusing, significantly extending the DOF of holographic image reconstruction. ASP: angular spectrum propagation. Modified from Ref. [129]

2.3.2 Results and Discussion

To demonstrate the success of HIDEF, in the initial set of experiments, I used aerosols that are captured by a soft impactor surface and imaged by an on-chip holographic microscope, where the optical field scattered by each aerosol interferes with the directly transmitted light forming an in-line hologram, sampled using a CMOS imager, without the use of any lenses[23]. The captured aerosols on the substrate are dispersed in multiple depths (z_2) as a result of varying particle mass, flow speed, and flow direction during the air sampling period[23]. Based on this set-up, the training image dataset had 176 digitally-cropped non-overlapping regions that only contained *particles located at the same depth*, which are further augmented by 4-fold to 704 regions by rotating them to 0, 90, 180 and 270 degrees. For each region, I used a single hologram intensity and back-propagated it to 81 random distances, spanning an axial range of -100 μm to 100 μm away from the correct global focus, determined by auto-focusing using the Tamura of the Gradient criterion[63]. I then used these complex-valued fields as the input to the network. The

target images used in the training phase (i.e., the reference images corresponding to the same samples) were reconstructed using multi-height phase-recovery (MH-PR) that utilized 8 different in-line holograms of the sample, captured at different z_2 distances, to iteratively recover the phase information of the sample, after an initial auto-focusing step performed for each height[193].

After this training phase, next I blindly tested the HIDEF network on samples that had *no overlap* with the training or validation sets; these samples contained *particles spread across different depths per image FOV*. **Figure 2.8** illustrates the success of HIDEF and how it simultaneously achieves an extended DOF and phase-recovery. For a given in-line hologram of the captured aerosols (**Figure 2.8(a)**), I first back-propagate the hologram intensity to a coarse distance of z_2 = 1 mm away from the active area of the CMOS imager, which is roughly determined based on the effective substrate thickness used in the experiment. This initial back-propagated hologram yields a strong twin image because of the short propagation distance (~1 mm) and the missing phase information. This complex-valued field, containing both the true and twin images, is then fed to the CNN. The output of the CNN is shown in **Figure 2.8(a)**, which demonstrates the extended DOF of HIDEF with various aerosols, spread over an axial range of ~90 μm, that are all brought into focus at the network output. In addition to bringing all the particles contained in a single hologram to a sharp focus, the network also performed phase-recovery, resulting in phase and amplitude images that are free from twin image and self-interference related artifacts. **Figure 2.8(b-c)** also compare the results of the network output with respect to a standard MH-PR approach that used eight in-line holograms to iteratively retrieve the phase information of the sample. These comparisons clearly demonstrate both the significantly extended DOF and phase-recovery performance of HIDEF, achieved using a single hologram intensity with a non-iterative inference time of < 0.2 s. In comparison, the iterative

MH-PR approach took ~4 s for phase-recovery and an additional ~2.4 s for auto-focusing to the individual objects at eight planes, totaling ~6.4 s for the same FOV and object volume, i.e., >30-fold slower compared to HIDEF.

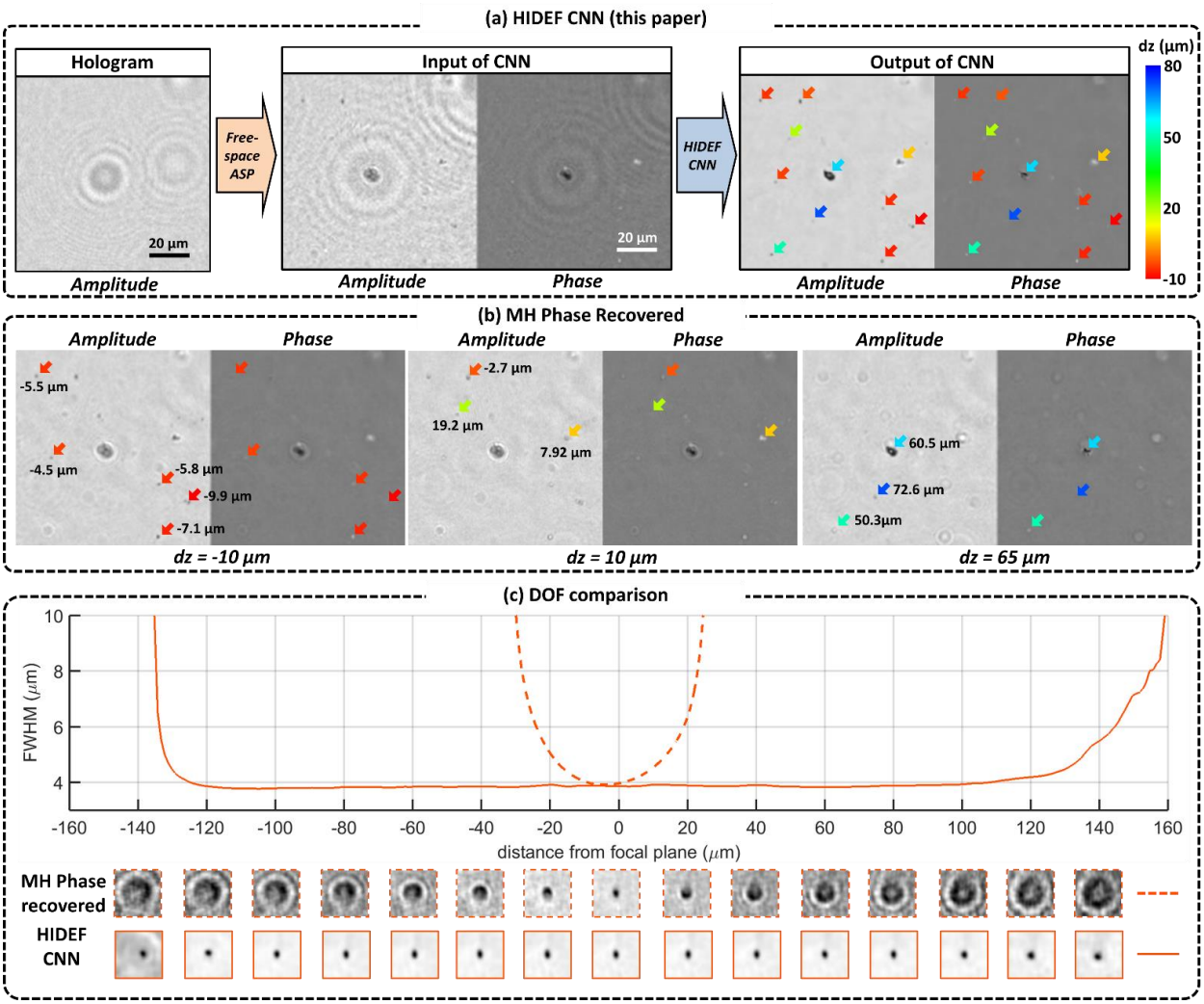

Figure 2.8 Extended-DOF reconstruction of aerosols at different depths using HIDEF. (a) After its training, HIDEF CNN brings all the particles within the FOV into focus, while also performing phase-recovery. Each particle's depth is color-coded, with respect to the back-propagation distance (1 mm), as shown with the color-bar on the right. (b) As a comparison, MH-PR images of the same FOV show that some of the particles come into focus at different depths, and become invisible or distorted at other depths. For each particle's arrow, the same color-coding is used as in (a). (c) The enhanced-DOF of HIDEF is illustrated by tracking a particle's amplitude full-width-half-maximum (FWHM) as a function of the axial defocus distance. HIDEF preserves the particle's FWHM

diameter and its correct image across a large DOF of >0.2 mm, which is expected since it was trained for this range of defocus (± 0.1 mm). On the other hand, MH-PR results show a much more limited DOF, as also confirmed with the same particle's amplitude images at different defocus distances, reported at the bottom. Reprinted from Ref. [129]

In these results, I used a coarse back-propagation step of 1 mm, before feeding the CNN with a complex-valued field. An important feature of HIDEF is that this back-propagation distance, z_2, does not need to be precise, and HIDEF can provide the same extended DOF image regardless of the initial back-propagation distance (z_2) selection initial, as long as it's within the range of defocusing that the network was trained for (+/- 0.1 mm in this case, see Ref. [129] for details).

Interestingly, although the network was only trained with *globally* de-focused hologram patches that only contain particles at the same depth/plane, it learned to individually focus various particles that lie at different depths within the same FOV (see **Figure 2.8(a)**). Based on this observation, one can argue that the HIDEF network does *not* perform the physical equivalent of free-space back-propagation of a certain hologram FOV to a focus plane. Instead, it statistically learns both in-focus and out-of-focus features of the input field, segments the out-of-focus parts and replaces them with in-focus features, in a parallel manner for a given hologram FOV. From an algorithm time-complexity perspective, this is a fixed processing time for a given hologram patch, i.e., a complexity of $O(1)$, instead of the conventional $O(nm)$, where n defines the number of individual object points or particles within the 3D sample volume, and m is the discrete focusing search space.

Based on the above argument, if the network statistically learns both in-focus and out-of-focus features of the sample, one could think that this approach should be limited to relatively sparse objects (such as the one shown in **Figure 2.8**). To test this hypothesis with non-sparse samples, next I tested HIDEF on the holograms of spatially connected objects such as tissue

slices, where there is no opening or empty region within the sample plane. For this goal, using the same CNN architecture, I trained the network with 1,119 hologram patches (corresponding to breast tissue sections used in histopathology), which were randomly propagated to 41 distances spanning an axial defocus range of -100 µm to 100 µm with respect to the focal plane. In this training phase, I used MH-PR images as the target/reference. The blind testing results are summarized in **Figure 2.9**, which clearly demonstrate that HIDEF can simultaneously perform both phase-recovery and auto-focusing for an arbitrary, non-sparse and connected sample.

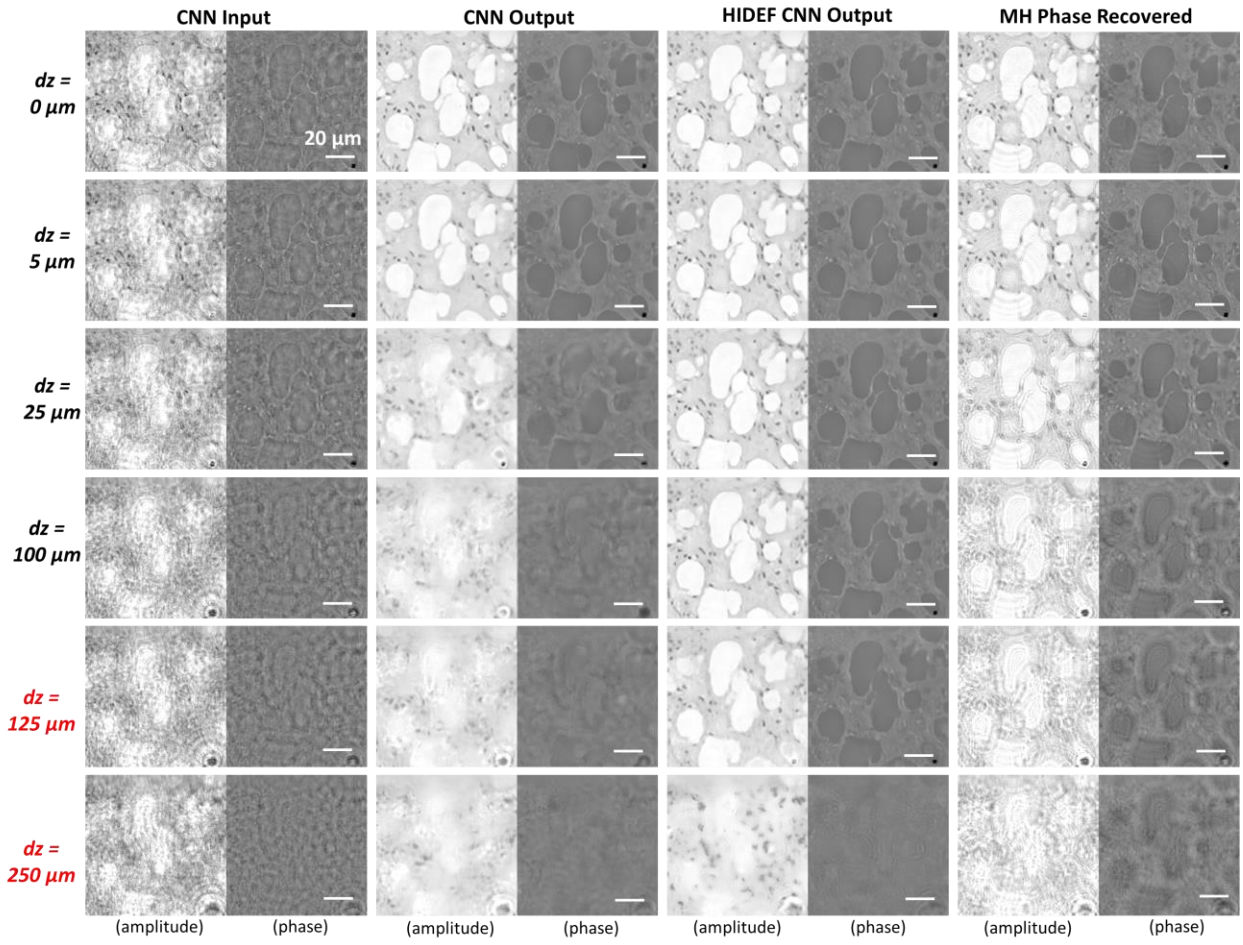

Figure 2.9 Comparison of HIDEF results against free-space back-propagation (CNN Input) and MH-PR (MH Phase Recovered) results, as a function of axial defocus distance (dz). The test sample is a thin section of a human breast tissue sample. The first two columns use a single intensity hologram, whereas the third column (MH-PR) uses eight in-line holograms of the same sample, acquired at different heights. These results clearly demonstrate

that the HIDEF network simultaneously performs phase-recovery and auto-focusing over the axial defocus range that it was trained for (i.e., $|dz| \leq 100$ μm in this case). Outside this training range (marked with red dz values), the network output is not reliable. See Visualization 3 and Visualization 4 for a detailed comparison. Scale bar: 20 μm. Reprinted from Ref. [129]

In **Figure 2.9**, we also see that MH-PR images naturally exhibit a limited DOF: even at an axial defocus of ~5 μm, some of the fine features at the tissue level are distorted. With more axial de-focus, the MH-PR results show significant artificial ripples and loss of further details. HIDEF, on the other hand, is very robust to axial defocusing, and is capable of correctly focusing the entire image and its fine features, while also rejecting the twin image artifact at different de-focus distances, up to the range that it was trained for (±0.1 mm).

However, as illustrated in **Figure 2.9**, beyond its training range, HIDEF starts to generate false features. Note that we do not see a physical smearing or diffraction-related smoothening effect as one continues to defocus in a range that the network was not trained for, rather in this defocus range that is "new" to the network, it still gives relatively sharp, but unrelated features. This implies that the network does not learn or generalize a specific physical process such as wave propagation, hologram formation or light interference; if it were to generalize such physical processes, one would not see sudden appearances of completely unrelated spatial features at the network output as one gradually goes outside the axial defocus range that it was trained for. In fact, this highlights a unique aspect of deep learning-based and data-driven holographic image reconstruction framework that is presented here: the output images of the network are driven by the image transformation that the neural network was trained for (between the input and gold standard label images), and this learned transformation creates output images that deviate from wave equation-based (physics-driven) solutions. An important example of this

is shown in **Figure 2.8**, where the network rapidly performed phase recovery and auto-focusing on all the particles that lie at different depths within the sample volume, bringing all the particles into focus at the output image. This image transformation from the network input to the output deviates from wave equation-based solutions that propagate a phase recovered field to different planes, where some particles will be in focus and some others will be out of focus, driven by free-space wave propagation.

To further quantify the improvements made by HIDEF, next I compared the amplitude of the network output image against the MH-PR result at the correct focus of the tissue section, and used the structural similarity (SSIM) index [181] for this comparison, defined as in equation (2.13). **Figure 2.10** shows the mean SSIM index calculated across an axial de-focus range of -100 μm to 100 μm, which was averaged across 180 different breast tissue FOVs that were blindly tested. Consistent with the qualitative comparison reported in **Figure 2.9**, HIDEF outputs SSIM values that are significantly higher than the hologram intensities back-propagated to the exact focus distances, owing to the phase-recovery capability of the network. Furthermore, as shown in **Figure 2.10**, compared to a CNN (with the same network architecture) that is trained using only in-focus holograms (with exact z_2 values), HIDEF has a much higher SSIM index for de-focused holograms, across a large DOF of ~0.2 mm. Interestingly, the network that is trained with in-focus holograms beats HIDEF for only one point in **Figure 2.10**, i.e., for dz = 0 μm, which is expected as this is what it was specifically trained for. However, this small difference in SSIM (0.78 vs. 0.76) is visually negligible (see Ref. [129] for details).

So far, I demonstrate the unique capabilities of HIDEF to simultaneously perform phase-recovery and auto-focusing, yielding at least an order of magnitude increase in the DOF of the reconstructed images, as also confirmed by **Figure 2.8** and **Figure 2.9**. To further extend the

DOF of the neural network output beyond 0.2 mm, one can use a larger network (with more layers, weights and biases) and/or more training data, containing severely defocused images as part of its learning phase. Certainly, the proof-of-concept DOF enhancement reported here is not an ultimate limit for the presented approach. In fact, to better emphasize this opportunity I also trained a third neural network, following the HIDEF architecture, with a training image set that contained randomly defocused holograms of breast tissue sections, with an axial defocus range of -0.2 mm to 0.2 mm. The performance comparison of this new network against the previous one (demonstrated in **Figure 2.9**) is reported in **Figure 2.10**. As shown in this comparison, by using a training image set that included more defocused holograms, I was able to significantly extend the axial defocus range to 0.4 mm (i.e., +/- 0.2 mm), where the HIDEF network successfully performed both auto-focusing and phase-recovery, at the same output image.

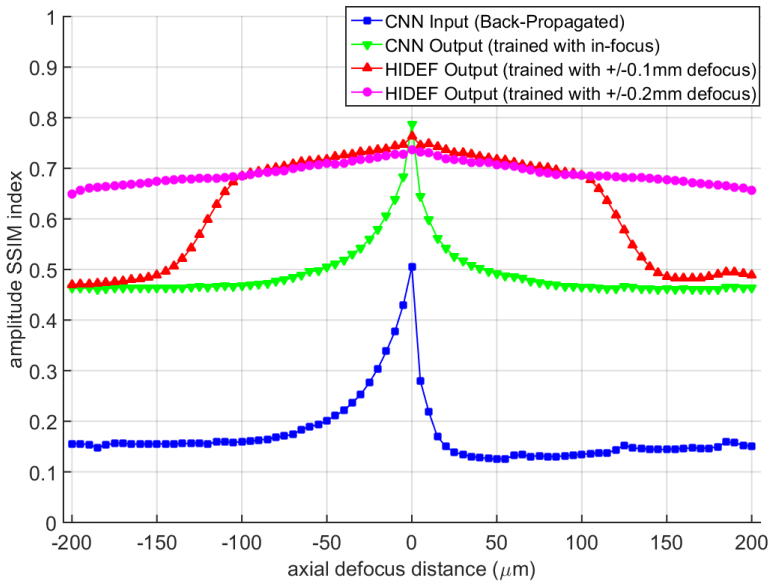

Figure 2.10 SSIM values as a function of the axial defocus distance. Each one of these SSIM curves is averaged over 180 test FOVs (512-by-512 pixels) corresponding to thin sections of a human breast tissue sample. The results confirm the extended-DOF of the HIDEF network output images, up to the axial defocus range that it was trained for. Reprinted from Ref. [129]

The results reported here provide compelling evidence for some unique opportunities created by statistical image reconstruction methods enabled by especially deep learning. HIDEF can be especially useful for wide-field, holographic imaging applications by digitally bringing a large sample volume into focus in real time, such as field-portable and cost-effective imaging flow-cytometers [194], enabling label-free and high-throughput screening of large volumes of liquid samples in field settings. As another example, this deep-learning based reconstruction method was also used to characterize particle aggregation-based bio-sensors over a wide imaging field of view greater than 20 mm^2 and achieved high throughput and rapid detection of viruses[195]. In yet another example, a portable and cost-effective device that senses bio-aerosols in the field was developed, which used HIDEF to rapidly reconstruct the microscopic images of captured bio-aerosols for their automatic detection and label-free classification, achieving an accuracy of >94% for different types of pollen and mold spores [24], which I will expand in section 3.2. In addition, the HIDEF approach can also be easily adapted to computationally extend the DOF of other imaging modalities, including e.g., fluorescence microscopy.

2.3.3 Methods

Network and architecture

The architecture of the CNN architecture is inspired by U-Net [196], and it consists of a down-sampling path as well as a symmetric up-sampling path (see section 2.2.2 and **Figure 2.6(a)** for details). Through a chain of down-sampling operations, the network learns to capture and separate the true image and twin image spatial features of a holographic input field at different scales[32]. This CNN architecture is implemented using TensorFlow, an open-source deep learning software package [110]. During the training phase, the CNN minimizes the l1-norm distance of the network output from the target/label images, and iteratively updates the network's weights and

biases using the adaptive moment estimation (Adam) optimizer [197], with a learning rate of 10^{-4}. For each image dataset, the ratio of the training to cross-validation was set to 14:3. The training and blind testing of the network were performed on a PC with six-core 3.60 GHz CPU, 16GB of RAM, using Nvidia GeForce GTX 1080Ti GPU. On average, the training process takes ~40 h for e.g., 200,000 iterations, corresponding to ~100 epochs. After the training, the network inference time for a hologram patch of 512 x 512 pixels (with phase and amplitude channels) is < 0.2 s.

2.4 Holography reconstruction with bright field contrast

Digital holographic microscopy enables the 3D reconstruction of volumetric samples from a single snapshot hologram. However, unlike a conventional bright-field microscopy image, the quality of holographic reconstructions is compromised by interference fringes as a result of twin images and out-of-plane objects. Here, I demonstrate that cross-modality deep learning using a generative adversarial network (GAN) can endow holographic images of a sample volume with bright-field microscopy contrast, combining the volumetric imaging capability of holography with the speckle- and artifact-free image contrast of incoherent bright-field microscopy. I illustrate the performance of this "*bright-field holography*" method through the snapshot imaging of bioaerosols distributed in 3D, matching the artifact-free image contrast and axial sectioning performance of a high-NA bright-field microscope. This data-driven deep-learning-based imaging method bridges the contrast gap between coherent and incoherent imaging and enables the snapshot 3D imaging of objects with bright-field contrast from a single hologram, benefiting from the wave-propagation framework of holography.

2.4.1 Introduction

Digital holographic microscopy encodes the volumetric information of a sample into a single 2D diffraction pattern. Thus, digital holographic microscopy enables the reconstruction of volumetric samples from a single hologram without mechanical scanning [16,34,62,183,184,187]. However, for most practical applications, holographic images cannot match the speckle- and artifact-free image contrast of incoherent bright-field microscopy. Some of these holographic artifacts include twin-image and self-interference artifacts, which are related to missing phase information; additional artifacts appear due to the long coherence length/diameter of the illumination source, which creates speckle and background interference from out-of-focus or unwanted objects in the optical path. Stated differently, since the point spread function (PSF) of a coherent imaging system has non-diminishing ripples along both the lateral and axial directions, out-of-focus objects will create interference fringes overlapping with the in-focus objects in the holographic reconstruction, which degrades the image contrast when reconstructing volumetric samples. These issues can be partially mitigated by using different holographic reconstruction methods and sometimes also by using additional measurements[17,32,34,71,76,91,127–129].

Here, I use a deep neural network to perform cross-modality image transformation from a digitally backpropagated hologram corresponding to a given depth within the sample volume into an image that is equivalent to a bright-field microscopy image acquired at the same depth (**Figure 2.11**). Since a single hologram is used to digitally propagate image information to different sections of the sample to virtually generate a bright-field equivalent image of each section, this approach combines the snapshot volumetric imaging capability of digital holography with the speckle- and artifact-free image contrast and axial sectioning performance of bright-field microscopy. Following its training, the deep neural network has learned the statistical image transformation between a holographic imaging system and an incoherent bright-field

microscope; therefore, I refer to this approach as “bright-field holography”. In some sense, deep learning brings together the best of both worlds by fusing the advantages of both the holographic and incoherent bright-field imaging modalities.

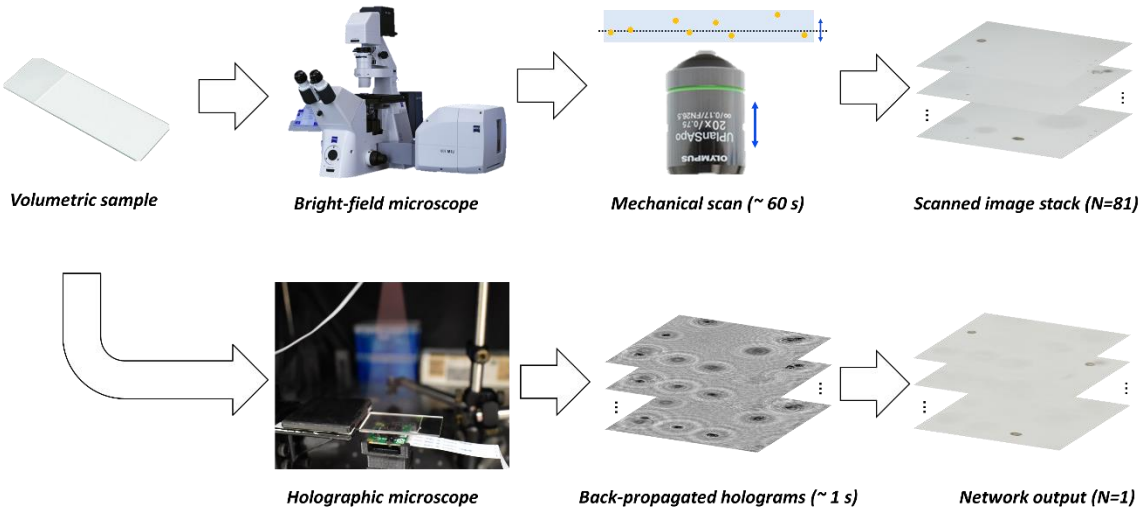

Figure 2.11 Bright-field holography. High-contrast bright-field imaging of a volumetric sample requires mechanical axial scanning and the acquisition of many successive images (e.g., N=101 here spans ±500 µm in depth). Bright-field holography, enabled by deep learning, fuses the volumetric imaging capability of holography with the speckle- and artifact-free image contrast of incoherent bright-field microscopy to generate bright-field equivalent images of a volume from a single hologram (N=1 image). Reprinted from Ref. [139]

2.4.2 Results and discussion

Holographic reconstruction with brightfield contrast and color

I used a generative adversarial network (GAN)[170,198,199] to perform the holographic to bright-field image transformation (**Figure 2.6**). The network’s training dataset was made up of images from pollen samples captured on a flat substrate using a sticky coverslip[24]. The coverslip was scanned in 3D using a bright-field microscope (Olympus IX83, 20×/0.75 NA objective lens), and a stack of 121 images with an axial spacing of 0.5 µm was captured for each region of interest to constitute the ground-truth labels. Next, using a lens-less holographic microscope[24], in-line

holograms were acquired corresponding to the same fields-of-view (FOV) scanned with the bright-field microscope. By progressively applying a series of image registration steps from the global coordinates to the local coordinates of each image patch, the backpropagated holograms at different depths were precisely matched to the bright-field microscopy ground-truth image stack in both the lateral and axial directions (see Methods for details). These registered pairs of backpropagated holograms and bright-field microscopy images were then cropped into ~6,000 patches of 256 × 256 pixels for training.

It should be emphasized that these steps need to be performed only once for the training of the GAN architecture, after which the generator network can blindly take a new hologram that it has never seen before and infer the corresponding bright-field image at any arbitrary depth within the sample volume in nearly real time (e.g., the inference time for a FOV of ~0.15 mm^2 is ~0.1 s using a single Nvidia 1080 Ti GPU). **Figure 2.12** presents an example of these blind testing results for several pollen mixtures, where the backpropagated holograms are compromised by twin-image and self-interference artifacts as well as speckle and out-of-focus interference. On the other hand, the generator network's output image for each depth clearly shows improved contrast and is free of the artifacts and noise features observed in the backpropagated holograms. These results match well with the corresponding bright-field images (the ground truth) at the same sample depths. In addition, the deep network correctly colorizes the output images based on the morphological features in the complex-valued input image, using an input hologram acquired with a monochrome sensor (Sony IMX219PQ, 1.12 μm pixel size) and narrowband illumination ($\lambda = 850\ nm$, bandwidth ~ 1 nm), such that the output matches the color distribution of the bright-field ground-truth image. This can be seen in **Figure 2.12** for the yellow ragweed pollens and oak tree pollens as well as the white Bermuda grass pollens.

Furthermore, the root mean square error (RMSE), structural similarity index (SSIM)[181], and universal image quality index (UIQI) [200] were used to quantitatively demonstrate the close resemblance of the network inferences to the bright-field microscopy ground-truth images, as shown in **Figure 2.12**.

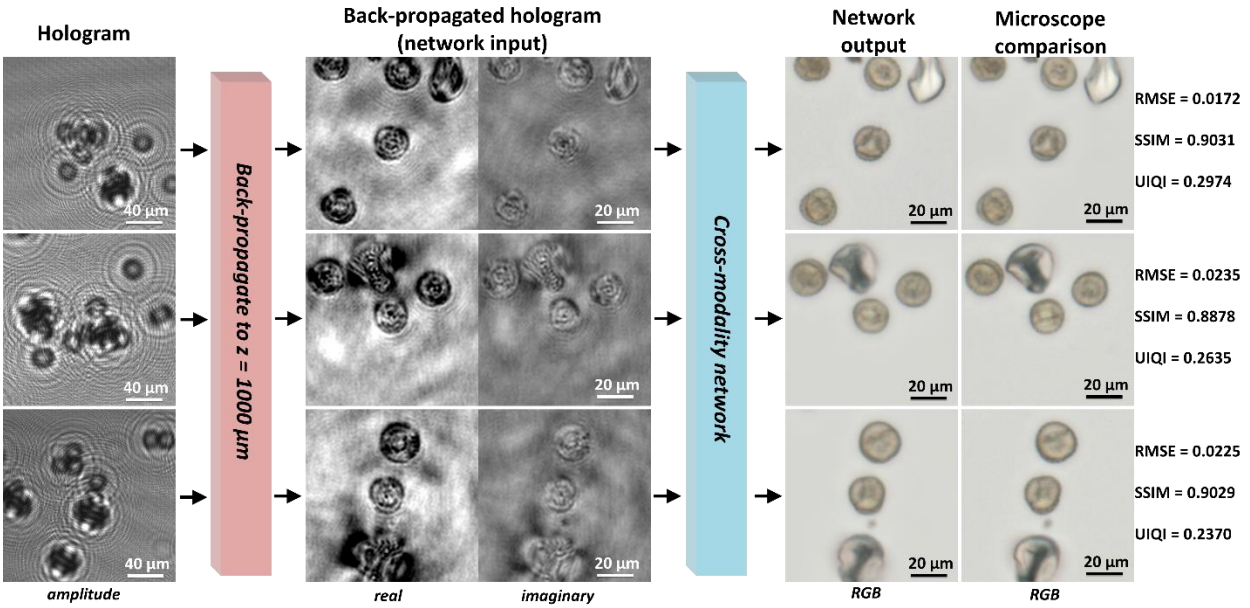

Figure 2.12 Imaging of a pollen mixture captured on a substrate. Each input hologram is shown with a larger FOV to better illustrate the fringes. Each network output image is quantitatively compared against the corresponding bright-field microscopy ground-truth image using the root mean square error (RMSE), the structural similarity index (SSIM), and the universal image quality index (UIQI). Reprinted from Ref. [139]

Blind inference of the network on volumetric 3D sample

Although the deep network was trained only with pollen mixtures captured on 2D substrates, it can successfully perform inference for the volumetric imaging of samples at different depths. **Figure 2.13** illustrate a pollen mixture captured in 3D in a bulk volume of polydimethylsiloxane (PDMS) with a thickness of ~800 μm. A single in-line hologram of this sample (**Figure 2.13(c)**) was captured and numerically backpropagated to different depths within the sample volume. By

feeding these backpropagated holographic images into the trained network, we obtained output images (**Figure 2.13**) that are free of speckle artifacts and various other interferometric artifacts observed in holography (e.g., twin images, fringes related to out-of-focus objects, and self-interference). These images match the contrast and depth-of-field (DOF) of bright-field microscopy images that were *mechanically* focused onto the same plane within the 3D sample.

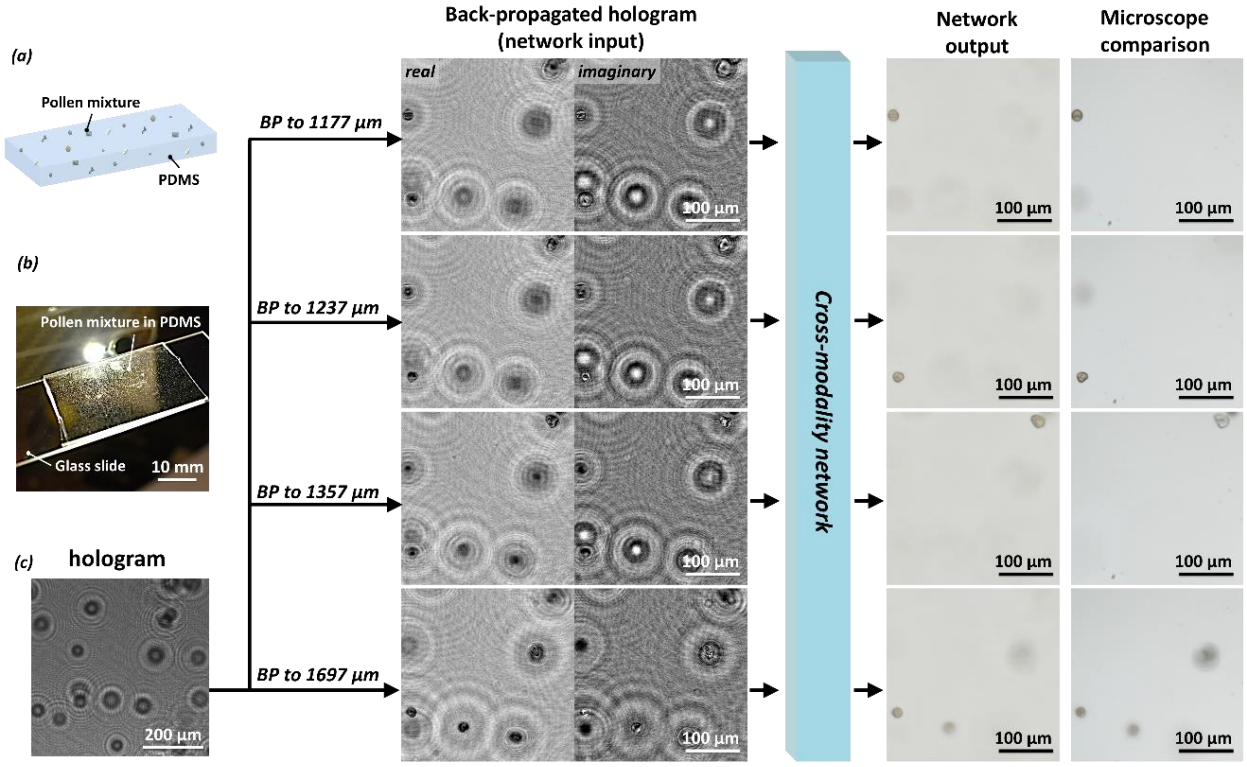

Figure 2.13 Use of cross-modality deep learning in bright-field holography to fuse the volumetric imaging capability of holography with the speckle- and artifact-free image contrast performance of incoherent bright-field microscopy. The pollen sample is dispersed in 3D throughout a bulk volume of PDMS (thickness ~ 800 µm). BP: digital backpropagation. Reprinted from Ref. [139]

It is also worth noting that the snapshot volumetric reconstruction performance presented in this work cannot be obtained through standard coherent denoising or phase recovery methods. To provide an example of this, in **Figure 2.14**, I compare the results of an object-support-based

phase recovery method [17,71] applied to the same sample hologram that was backpropagated to different heights. As shown in this figure, the iterative phase recovery method indeed improved the contrast-to-noise ratio (CNR)[201] of the backpropagated holographic images from ~2 to ~3, especially suppressing some of the twin-image-related artifacts. However, the out-of-focus fringes created by the 3D object were not adequately sectioned out and remained as reconstruction artifacts even after iterative phase recovery. In contrast, the deep neural network output transformed the defocused coherent fringes into diminished incoherent blobs, achieving a high CNR of >15-25, very well matching the ground-truth images captured by the high-NA bright-field microscope, as shown in **Figure 2.14**.

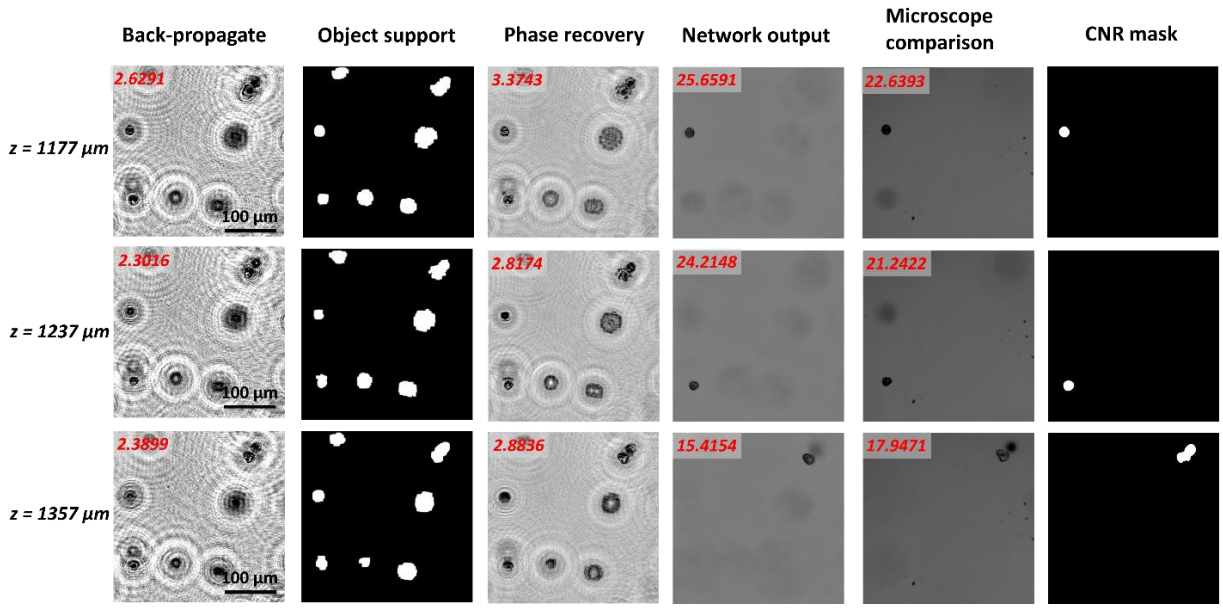

Figure 2.14 Comparison of bright-field holography against an iterative phase recovery method for a volumetric pollen sample. Object support-based phase recovery[17,71] was used (with 20 iterations). The network output and bright-field microscope ground truth images were converted into gray scale for comparison. Contrast to noise ratio (CNR)[201], which is defined as the ratio of the contrast (the average of the pixel amplitudes outside the CNR mask minus the average of the pixel amplitudes inside the mask) with respect to noise (the standard deviation of the pixel amplitudes outside the mask), was used to quantitatively compare the results, with the corresponding CNR values marked by a red number on the upper-left corner of each image. Reprinted from Ref. [139]

Transforming holographic PSF to incoherent bright field PSF

To further quantify this cross-modality transformation performance, I imaged samples containing 1 μm polystyrene beads and trained another GAN following the same method. Next, I blindly tested a sample containing 245 individual/isolated microbeads and measured their 3D PSF distributions before and after GAN inference (**Figure 2.15**). An example of this comparison is shown in **Figure 2.15(a)**, where the backpropagated holograms contain significant interference artifacts that were removed by the GAN, yielding images that match the high contrast of the mechanically scanned bright-field microscopy ground-truth images.

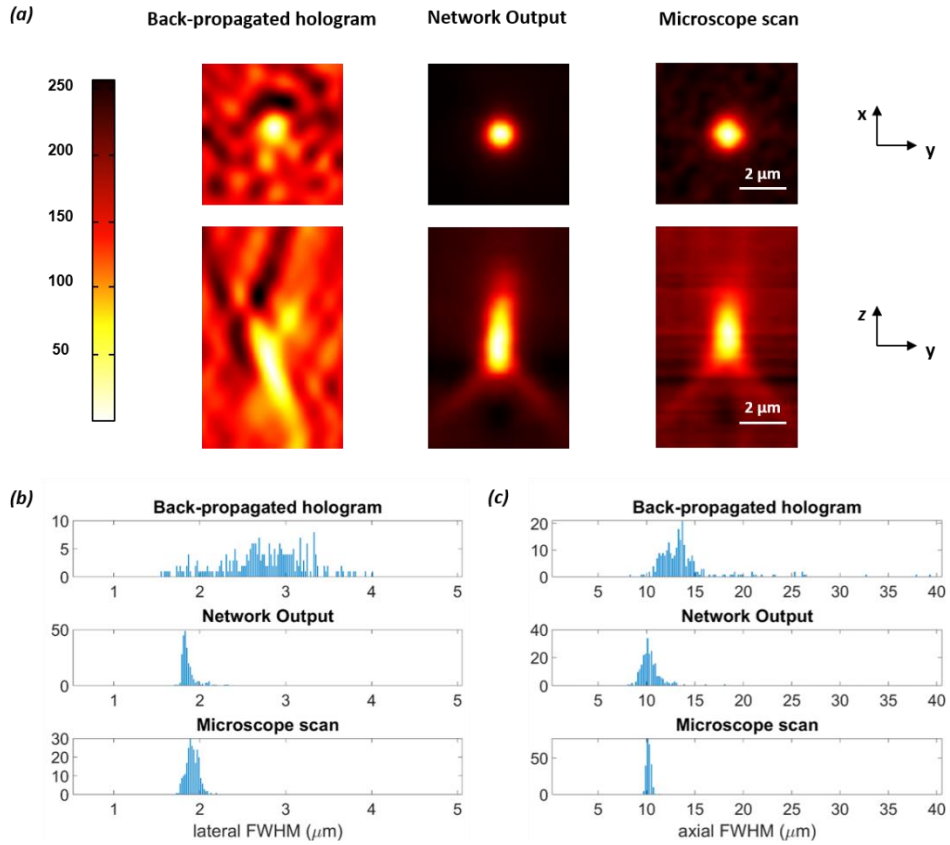

Figure 2.15 3D PSF comparison using 1 μm beads. (a) 3D imaging of a single microbead and a comparison of the standard holographic backpropagation results against the network output and the images captured by a scanning bright-field microscope via N=81 scans with an axial step size of 0.5 μm. (b) Lateral PSF FWHM histogram comparison corresponding to 245 individual/isolated microbeads. (c) Same as in (b), except for the axial PSF FWHM histograms. Reprinted from Ref. [139]

Figure 2.15(b) shows the distributions of the lateral and axial full-width-at-half-maximum (FWHM) values corresponding to the 3D PSFs obtained using these 245 microbeads. Due to the interference artifacts and low contrast, the FWHM values of the PSFs of the backpropagated hologram (input) are randomly distributed in the lateral direction, with a median FWHM of 2.7176 μm. In contrast, the lateral FWHM values of the PSFs of the GAN output images are monodisperse, with a median FWHM of 1.8254 μm, matching that of the scanning bright-field microscopy ground truth (1.8719 μm). Due to the longer coherence length, the PSFs of the backpropagated hologram (input) are longer in the axial direction, with a median FWHM of 12.9218 μm, compared to the scanning bright-field microscopy ground truth, with a median FWHM of 9.8003 μm. The network inference results show a significantly narrower PSF distribution in the axial direction, with a median FWHM of 9.7978 μm, very well matching that of the ground truth obtained with the scanning bright-field microscope (**Figure 2.15(b)**).

2.4.3 Conclusion

This deep-learning-enabled cross-modality image transformation between holography and bright-field imaging can potentially eliminate the need to mechanically scan a volumetric sample. It is benefited from the digital wave-propagation framework of holography to virtually scan throughout the sample volume, and each one of these digitally propagated fields is transformed into bright-field microscopy equivalent images that exhibit the spatial and color contrast as well as the shallow DOF expected from incoherent microscopy. In this regard, my deep-learning-enabled hologram transformation network achieves the best of both worlds by fusing the volumetric digital imaging capability of holography with the speckle- and artifact-free image contrast of bright-field microscopy. This capability can be especially useful for the rapid volumetric imaging of samples flowing within a liquid [194]. This approach can also be applied to

other holographic microscopy and/or incoherent microscopy modalities to establish a statistical image transformation from one mode of coherent imaging into another incoherent microscopy modality. This work enables the inference of a whole 3D sample volume from a single snapshot hologram, thus reintroducing coherent holographic imaging as a powerful alternative to high-NA bright-field microscopy for the task of high-throughput volumetric imaging, and therefore represents a unique contribution to the field of coherent microscopy.

2.4.4 Methods

Digital holographic and scanning bright-field microscopy image acquisition

The holographic images were acquired using a customized lens-less imaging system. The system consisted of a vertical-cavity surface-emitting laser (VCSEL) diode ($\lambda = 850\,nm$) for illumination, a complementary metal–oxide–semiconductor (CMOS) image sensor (Sony IMX219PQ, 1.12 μm pixel size) and a Raspberry Pi 2 for system control. This near-infrared illumination was chosen to enable the use of all four Bayer channels of the color image sensor chip to improve the pixel-size-limited resolution of the hologram that could be achieved in a single snapshot. The sample was mounted on a 3D printed sample holder placed ~500 μm above the image sensor surface. The illumination source was placed ~8 cm above the sample plane without any additional spatial or spectral filter.

The bright-field microscopy images were captured by an inverted scanning microscope (IX83, Olympus Life Science) using a 20×/0.75 NA objective lens. The microscope scanned each sample at different lateral locations, and at each location, an image stack of -30 μm to 30 μm with a 0.5 μm step size was captured. After the capture of these bright-field images, the microscopy image stack was aligned using the ImageJ plugin StackReg[202], which corrected the rigid shift and rotation caused by the inaccuracy of the microscope scanning stage.

Image registration and training data preparation

The success of the cross-modality transform relies on accurate registration of the back-propagated holograms with the scanning bright-field microscope images in 3D. This registration can be divided into two parts. The first part laterally matches a bright-field image (2048 × 2048 pixels) to that of the hologram using SURF feature based registration [203]. The second part further refines the registration in x-y and z directions, where Pyramid elastic registration[204] was performed on the image pairs closest to the focal plane to further warp the image pairs to each other. To match the heights of the backpropagated holograms and the corresponding bright-field microscopy image stacks, the focal planes were estimated and cross-registered as "zero" height, and the relative axial propagation distance was determined to match the axial scanning step size of the bright-field microscope (0.5 μm). The digital hologram's focal plane was estimated using the holographic autofocusing with edge sparsity criterion[63]. (see Ref.[139] for details).

Network and training

The GAN implemented here consisted of a generator network and a discriminator network, as shown in **Figure 2.6**. The generator network employed a variation of the U-Net[196] design with minor modifications and additional residual connections[103]. The discriminator network was a convolutional neural network with six convolutional blocks and two fully connected (linear) layers. The original training data consisted of ~6,000 image pairs, which were augmented to 30,000 image pairs by random rotation and flipping. The validation data were not augmented. In each training iteration, the generator network was updated six times using Adam optimizer with learning rate 10^{-4}, then the discriminator network was updated three times using Adam optimizer with learning rate 3×10^{-5}. The network was evaluated on validation set every 50 iterations for the mean absolute error (MAE) loss, and the best one was chosen with the lowest MAE loss. The

network was built on TensorFlow[110]. The training and inference were performed on a PC with a six-core 3.6 GHz CPU and 16 GB of RAM using a Nvidia GeForce GTX 1080 Ti GPU. On average, the training process took ~ 90 hours for ~ 50,000 iterations (equivalent to ~ 40 epochs). After training, the network inference time was ~ 0.1 s for an image patch of 256×256 pixels.

2.5 Three-dimensional (3D) virtual refocusing of fluorescence images

Unlike holography, fluorescence microscopy lacks an image propagation and time-reversal framework, which necessitates scanning of fluorescent objects to obtain 3D images. In this section, I demonstrate that a deep convolutional neural network can be trained to virtually propagate and refocus a 2D fluorescence image onto user-defined 3D surfaces within the sample volume. With this data-driven computational microscopy framework, I imaged the neuron activity of a *Caenorhabditis elegans* worm in 3D at up to 100 Hz using a time-sequence of fluorescence images acquired at a single focal plane, digitally increasing the depth-of-field of the microscope by 20-fold without any axial scanning, additional hardware, or a trade-off of imaging resolution or speed. Furthermore, I demonstrate that this learning-based approach can correct for sample drift, tilt, and other image aberrations, all digitally performed after the acquisition of a single fluorescence image. This unique framework also cross-connects different imaging modalities to each other, enabling 3D refocusing of a single wide-field fluorescence image to match confocal microscopy images acquired at different sample planes. This deep learning-based 3D image refocusing method might be transformative for imaging and tracking of 3D fluorescent samples, especially over extended periods of time, mitigating photo-toxicity, sample drift, aberration and defocusing related challenges associated with standard 3D fluorescence microscopy techniques.

2.5.1 Introduction

Three-dimensional (3D) fluorescence microscopic imaging is essential for biomedical and physical sciences as well as engineering, covering various applications [205–211]. Unlike coherent imaging systems, such as holography, fluorescence emission from samples is both spatially and temporally incoherent, and therefore fluorescence microscopy in general lacks digital image propagation and time-reversal framework that is commonly employed in coherent microscopy for 3D imaging [212–214]. In fact, high-throughput acquisition of fluorescence image data for a 3D sample remains a challenge in microscopy research. 3D fluorescence information is usually acquired through scanning across the sample volume, where several 2D fluorescence images/measurements are obtained, one for each focal plane or point in 3D , which forms the basis of e.g., confocal [215], two-photon [209], light-sheet [211,216,217] or various super-resolution [206,208,218–221] microscopy techniques. However, because scanning is used, the image acquisition speed and the throughput of the system for volumetric samples are limited to a fraction of the frame-rate of the camera/detector, even with optimized scanning strategies [215] or PSF engineering [211,222]. Moreover, because the images at different sample planes/points are not acquired simultaneously, the temporal variations of the sample fluorescence can inevitably cause image artifacts. Another concern is the photo-toxicity of illumination and photo-bleaching of fluorescence since parts of the sample can be repeatedly excited during the scanning process.

To overcome some of these challenges, non-scanning 3D fluorescence microscopy methods have also been developed, so that the entire 3D volume of the sample can be imaged at the same speed as the detector framerate. One of these methods is fluorescence light-field microscopy [210,223–226]. This system typically uses an additional micro-lens array to encode the 2D angular information as well as the 2D spatial information of the sample light rays into image sensor pixels; then a 3D focal stack can be digitally reconstructed from this recorded 4D light-field.

However, using a micro-lens array reduces the spatial sampling rate which results in a sacrifice of both the lateral and axial resolution of the microscope. Although the image resolution can be improved by 3D deconvolution [210] or compressive sensing [224] techniques, the success of these methods depends on various assumptions regarding the sample and the forward model of the image formation process. Furthermore, these computational approaches are relatively time-consuming as they involve an iterative hyperparameter tuning as part of the image reconstruction process. A related method termed multi-focal microscopy has also been developed to map the depth information of the sample onto different parallel locations within a single image [227,228]. However, the improved 3D imaging speed of this method also comes at the cost of reduced imaging resolution or field-of-view (FOV) and can only infer an experimentally pre-defined (fixed) set of focal planes within the sample volume. As another alternative, the fluorescence signal can also be optically correlated to form a Fresnel correlation hologram, encoding the 3D sample information in interference patterns [5,10,229]. To retrieve the missing phase information, this computational approach requires multiple images to be captured for volumetric imaging of a sample. Quite importantly, all these methods summarized above, and many others, require the addition of customized optical components and hardware into a standard fluorescence microscope, potentially needing extensive alignment and calibration procedures, which not only increase the cost and complexity of the optical set-up, but also cause potential aberrations and reduced photon-efficiency for the fluorescence signal.

Here I introduce a digital image refocusing framework in fluorescence microscopy by training a deep neural network using microscopic image data, enabling 3D imaging of fluorescent samples using a single 2D wide-field image, without the need for any mechanical scanning, additional hardware or parameter estimation. This data-driven fluorescence image

refocusing framework does not need a physical model of the imaging system, and rapidly refocuses a 2D fluorescence image onto user-defined 3D surfaces. In addition to rapid 3D imaging of a fluorescent sample volume, it can also be used to digitally correct for various aberrations due to the sample and/or the optical system. This deep learning-based approach, termed as *Deep-Z*, computationally refocus a single 2D wide-field fluorescence image onto 3D surfaces within the sample volume, without sacrificing the imaging speed, spatial resolution, field-of-view, or throughput of a standard fluorescence microscope.

In *Deep-Z* framework, an input 2D fluorescence image (to be digitally refocused onto a 3D surface within the sample volume) is first appended with a user-defined digital propagation matrix (DPM) that represents, pixel-by-pixel, the axial distance of the target surface from the plane of the input image (**Figure 2.16**). *Deep-Z* is trained using a conditional generative adversarial neural network (GAN) [170,198] using accurately matched pairs of (1) various fluorescence images axially-focused at different depths and appended with different DPMs, and (2) the corresponding fluorescence images (i.e., the ground truth labels) captured at the correct/target focus plane defined by the corresponding DPM. Through this training process that only uses experimental image data without any assumptions or physical models, the generator network of GAN *learns to interpret the values of each DPM pixel as an axial refocusing distance*, and outputs an equivalent fluorescence image that is digitally refocused within the sample volume to the 3D surface defined by the user, where some parts of the sample are focused, while some other parts get out-of-focus, according to their true axial positions with respect to the target surface.

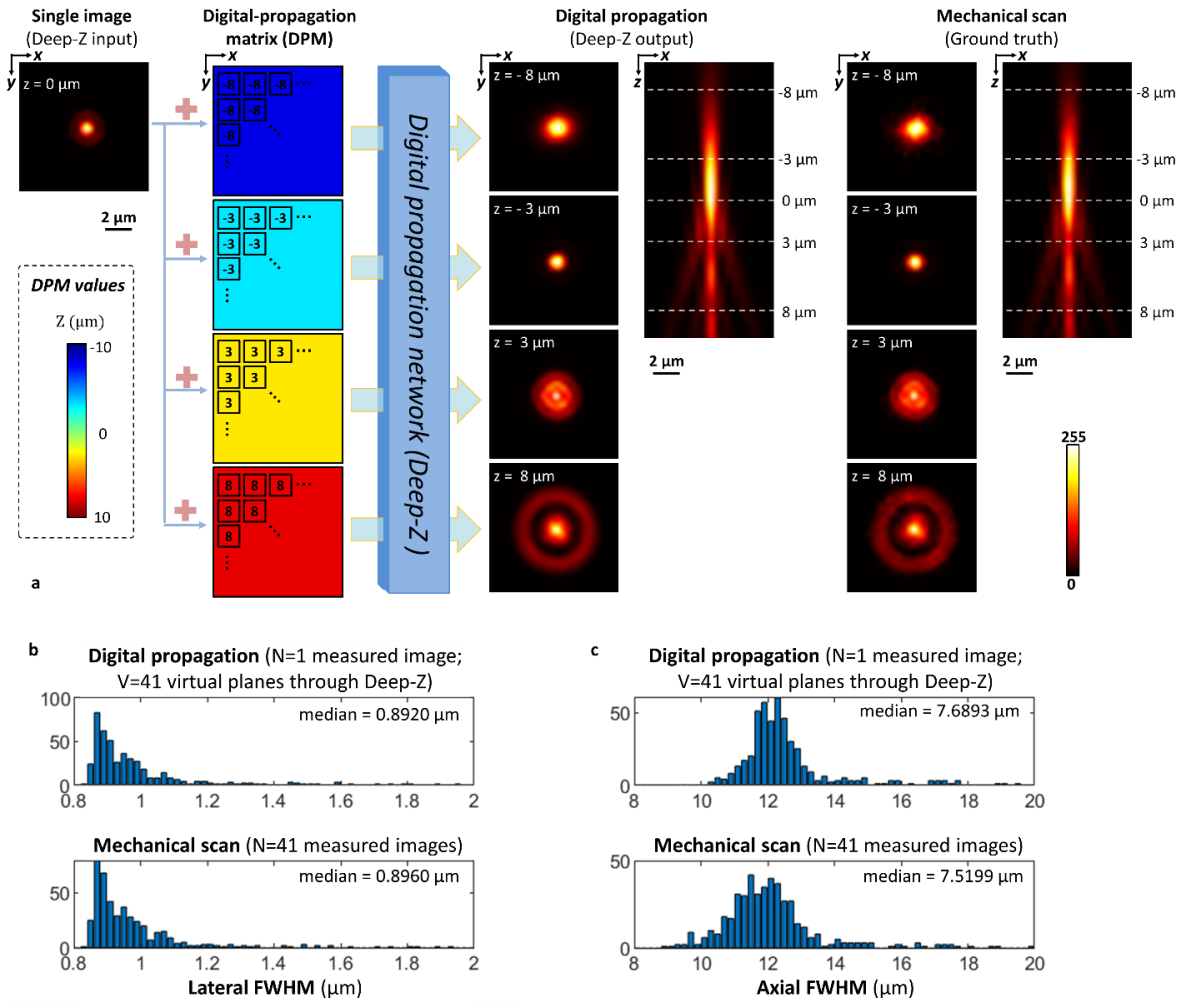

Figure 2.16 Refocusing of fluorescence images using *Deep-Z*. (a) By concatenating a digital propagation matrix (DPM) to a single fluorescence image, and running the resulting image through a trained *Deep-Z* network, digitally refocused images at different planes can be rapidly obtained, as if an axial scan is performed at the corresponding planes within the sample volume. The DPM has the same size as the input image and its entries represent the axial propagation distance for each pixel and can also be spatially non-uniform. The results of *Deep-Z* inference are compared against the images of an axial-scanning fluorescence microscope for the same fluorescent bead (300 nm), providing a very good match. (b) Lateral FWHM histograms for 461 individual/isolated fluorescence nano-beads (300 nm) measured using *Deep-Z* inference (N=1 captured image) and the images obtained using mechanical axial scanning (N=41 captured images) provide a very good match to each other. (c) Same as in (b), except for the *axial* FWHM measurements for the same data set, also revealing a very good match between *Deep-Z* inference results and the axial mechanical scanning results. Reprinted from Ref. [123].

To demonstrate the success of this unique fluorescence digital refocusing framework, I imaged *Caenorhabditis elegans* (*C. elegans*) neurons using a standard wide-field fluorescence microscope with a 20×/0.75 numerical aperture (NA) objective lens, and extended the native depth-of-field (DOF) of this objective (~1 μm) by ~20-fold, where a single 2D fluorescence image was axially refocused using *Deep-Z* to $\Delta z = \pm 10$ μm with respect to its focus plane, providing a very good match to the fluorescence images acquired by mechanically scanning the sample within the same axial range. Similar results were also obtained using a higher NA objective lens (40×/1.3 NA). Using this deep learning-based fluorescence image refocusing technique, I further demonstrated 3D tracking of the neuron activity of a *C. elegans* worm over an extended DOF of ±10 μm using a fluorescence video acquired at a *single* focal plane.

Furthermore, to highlight some of the additional degrees-of-freedom enabled by *Deep-Z*, I used spatially *non-uniform* DPMs to refocus a 2D input fluorescence image onto user-defined 3D surfaces to computationally correct for aberrations such as sample drift, tilt and spherical aberrations, all performed *after* the fluorescence image acquisition and without any modifications to the optical hardware of a standard wide-field fluorescence microscope.

Another important feature of this *Deep-Z* framework is that it permits *cross-modality digital refocusing* of fluorescence images, where the GAN is trained with gold standard label images obtained by a *different* fluorescence microscopy modality to teach the generator network to refocus an input image onto another plane within the sample volume, but this time to match the image of the same plane that is acquired by a different fluorescence imaging modality compared to the input image. I term this related framework *Deep-Z+*. To demonstrate this unique capability, I trained *Deep-Z+* with input and label images that were acquired with a wide-field fluorescence microscope and a *confocal* microscope, respectively, to blindly generate at the output of this

cross-modality *Deep-Z+*, digitally refocused images of an input wide-field fluorescence image that match confocal microscopy images of the same sample sections.

After its training, *Deep-Z* remains fixed, while the appended DPM provides a "depth tuning knob" for the user to refocus a single 2D fluorescence image onto 3D surfaces and output the desired digitally-refocused fluorescence image in a rapid non-iterative fashion. In addition to fluorescence microscopy, *Deep-Z* framework might potentially be applied to other incoherent imaging modalities, and in fact it bridges the gap between coherent and incoherent microscopes by enabling 3D digital refocusing of a sample volume using a single 2D incoherent image. *Deep-Z* is further unique in that it enables a computational framework for rapid transformation of a 3D surface onto another 3D surface within the fluorescent sample volume using a single forward-pass operation of the generator network.

2.5.2 Results

Digital refocusing of fluorescence images using *Deep-Z*

Deep-Z enables a single intensity-only wide-field fluorescence image to be digitally refocused to a user-defined surface within the axial range of its training. **Figure 2.16(a)** demonstrates this concept by digitally propagating a single fluorescence image of a 300 nm fluorescent bead (excitation/emission: 538 nm/584 nm) to multiple user defined planes. The native DOF of the input fluorescence image, defined by the NA of the objective lens (20×/0.75 NA), is ~ 1 μm; *Deep-Z* digitally refocus the image of this fluorescent bead over an axial range of ~ ±10 μm, matching the mechanically-scanned corresponding images of the same region of interest (ROI), which form the ground truth. Note that the PSF in **Figure 2.16(a)** is asymmetric in the axial direction, which provides *directional* cues to the neural network regarding the digital propagation of an input image by *Deep-Z*. Unlike a symmetric Gaussian beam[230], such PSF

asymmetry along the axial direction is ubiquitous in fluorescence microscopy systems[231]. To further quantify *Deep-Z* output performance I used PSF analysis; **Figure 2.16(b-c)** illustrate the histograms of both the *lateral* and the *axial* full-width-half-maximum (FWHM) values of 461 individual/isolated nano-beads distributed over ~ 500 × 500 μm^2. The statistics of these histograms very well agree with each other, confirming the match between *Deep-Z* output images calculated from a single fluorescence image (N=1 measured image) and the corresponding axially-scanned ground truth images (N=41 measured images). This quantitative match highlights the fact that *Deep-Z* indirectly learned, through image data, the 3D propagation of fluorescence light. However, this learned capability is limited to be within the axial range determined by the training dataset (e.g., ±10 μm in this work).

Next, I tested the *Deep-Z* framework by imaging the neurons of a *C. elegans* nematode expressing pan-neuronal tagRFP [232]. **Figure 2.17** demonstrates the blind testing results for *Deep-Z* based refocusing of different parts of a *C. elegans* worm from a *single* wide-field fluorescence input image. Using *Deep-Z*, non-distinguishable fluorescent neurons in the input image were brought into focus at different depths, while some other in-focus neurons at the input image got out-of-focus and smeared into the background, according to their true axial positions in 3D; see the cross-sectional comparisons to the ground truth mechanical scans provided in **Figure 2.17**. For optimal performance, this *Deep-Z* was specifically trained using *C. elegans* samples, to accurately learn the 3D PSF information together with the refractive properties of the nematode body and the surrounding medium. Similar results were also obtained for a 40×/1.3NA objective lens, where *Deep-Z* successfully refocused the input image over an axial range of ~ ±4 μm. In addition to 3D imaging of the neurons of a nematode, *Deep-Z* also works well to digitally refocus the images of fluorescent samples that are spatially denser such as the mitochondria and

F-actin structures within bovine pulmonary artery endothelial cells (BPAEC); see Ref.[123] for examples of these results.

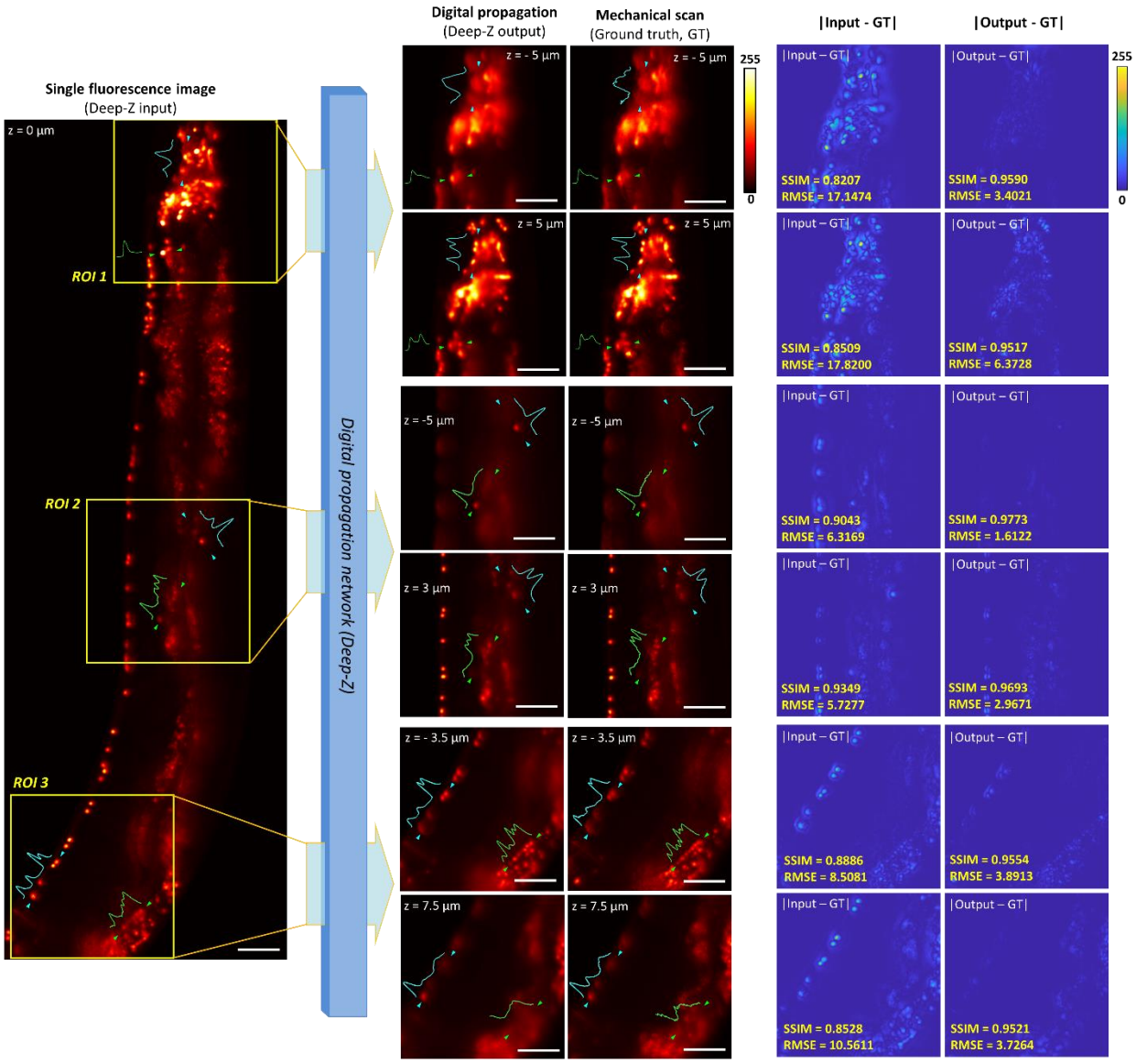

Figure 2.17 3D imaging of *C. Elegans* neuron nuclei using *Deep-Z*. Different ROIs are digitally refocused using *Deep-Z* to different planes within the sample volume; the resulting images provide a very good match to the corresponding ground truth images, acquired using a scanning fluorescence microscope. The absolute difference images of the input and output with respect to the corresponding ground truth image are also provided on the right, with structural similarity index (SSIM) and root mean square error (RMSE) values reported, further demonstrating the success of *Deep-Z*. Scale bar: 25 µm. Reprinted from Ref. [123].

3D functional imaging of *C. elegans* using *Deep-Z*

An important application of 3D fluorescence imaging is neuron activity tracking; for example genetically modified animals that express different fluorescence proteins are routinely imaged using a fluorescence microscope to reveal their neuron activity. To highlight the utility of *Deep-Z* for tracking the activity of neurons in 3D, I recorded a fluorescence video of a *C. elegans* worm at a single focal plane (z = 0 μm) for ~35 sec, using a 20×/0.8NA objective lens with two fluorescence channels: FITC for neuron activity and Texas Red for neuron locations. The input video frames were registered with respect to each other to correct for the slight body motion of the worm between the consecutive frames. Then, each frame at each channel of the acquired video were digitally refocused using *Deep-Z* to a series of axial planes from -10 μm to 10 μm with 0.5 μm step size, generating a virtual 3D fluorescence stack for each acquired frame. To quantify the neuron activity using *Deep-Z* output images, I segmented voxels of each individual neuron using the Texas Red channel (neuron locations), and tracked the change of the fluorescence intensity, i.e., $\Delta F(t) = F(t) - F_0$, in the FITC channel (neuron activity) inside each neuron segment over time, where $F(t)$ is the neuron fluorescence emission intensity and F_0 is its time average; see Methods for details. A total of 155 individual neurons in 3D were isolated using *Deep-Z* output images, as shown in **Figure 2.18(b)**, where the color represents the depth (z location) of each neuron. For comparison, the results of the same segmentation algorithm applied on just the input 2D image identified only 95 neurons without any depth information. **Figure 2.18(c)** plots the activities of the 70 most active neurons, which were grouped into clusters C1-C3 based on their calcium activity pattern similarities (see Methods for details). **Figure 2.18(c)** reports that cluster C3 calcium activities increased at t = 14 s, whereas the activities of cluster C2 decreased at a similar time point. These neurons very likely correspond to the motor neurons

type A and B that promote backward and forward motion, respectively, which typically anti-correlate with each other[233]. Cluster C1 features two cells that were comparatively larger in size, located in the middle of the worm. These cells had three synchronized short spikes at t = 4, 17 and 32 sec. Their 3D positions and calcium activity pattern regularity suggest that they are neuronal or muscle cells of the defecation system that initiates defecation in regular intervals in coordination with the locomotion system [234].

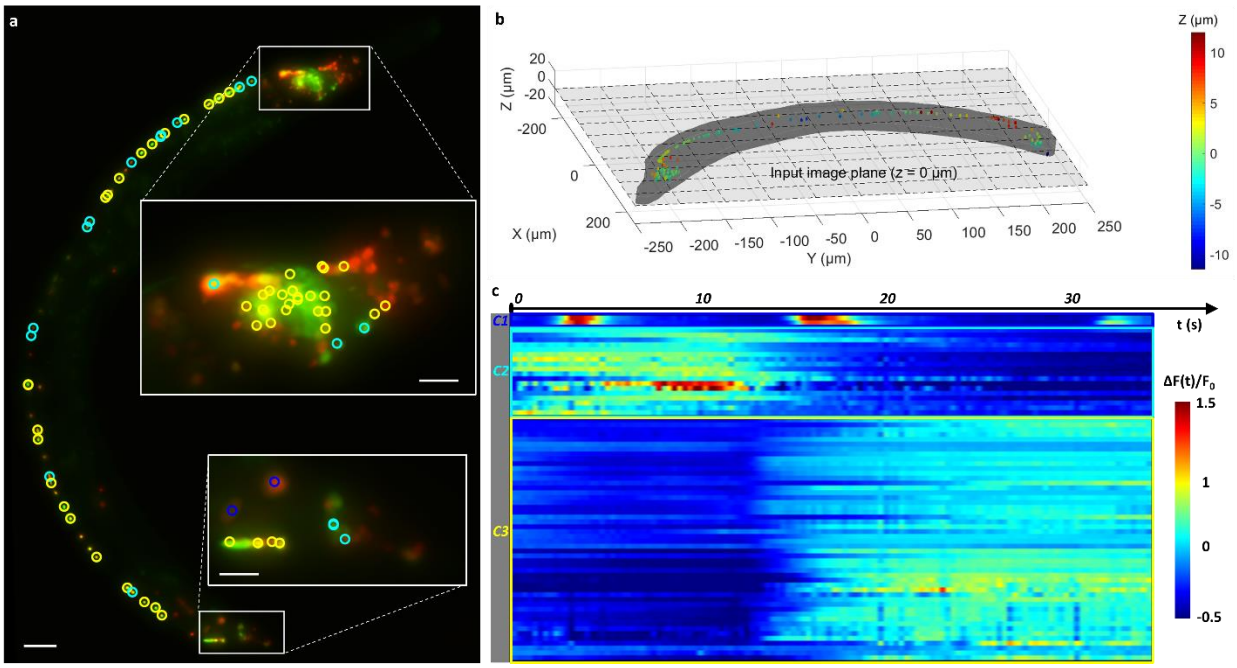

Figure 2.18 *C. Elegans* neuron activity tracking in 3D using *Deep-Z*. (a) Maximum intensity projection (MIP) along the axial direction of the median intensity image taken across the time sequence. The red channel (Texas red) labels neuron nuclei. The green channel (FITC) labels neuron calcium activity. A total of 155 neurons were identified, 70 of which were active in calcium activity. Scale bar: 25 μm. Scale bar for the zoom-in regions: 10 μm. (b) All the 155 localized neurons are shown in 3D, where depths are color-coded. (c) 3D tracking of neuron calcium activity events corresponding to the 70 active neurons. The neurons were grouped into 3 clusters (C1-C3) based on their calcium activity pattern similarity (see Methods section). The locations of these neurons are marked by the circles in (a). The colors of the circles represent different clusters: C1(blue), C2(cyan) and C3(yellow).

Note that all this 3D tracked neuron activity was in fact embedded in the input 2D fluorescence image sequence acquired at a single focal plane within the sample, but could not be readily inferred from it. Through the *Deep-Z* framework and its 3D refocusing capability to user defined surfaces within the sample volume, the neuron locations and activities were accurately tracked using a 2D microscopic time sequence, without the need for mechanical scanning, additional hardware, or a trade-off of resolution or imaging speed. Since *Deep-Z* generates temporally synchronized virtual image stacks through purely *digital* refocusing, it can be used to match the imaging speed to the limit of the camera framerate, by using e.g., the stream mode, which typically enables a short video of up to 100 frames per second. To highlight this opportunity, I used the stream mode of the camera of a Leica SP8 microscope (see the Methods section for details) and captured two videos at 100 fps for monitoring the neuron nuclei (under the Texas Red channel) and the neuron calcium activity (under the FITC channel) of a moving *C. elegans* over a period of 10 sec, and used *Deep-Z* to generate virtually refocused videos from these frames over an axial depth range of +/- 10 μm; see Ref. [123] for details.

***Deep-Z* based aberration correction using spatially non-uniform DPMs**

The results reported up to now used *uniform* DPMs in both the training phase and the blind testing in order to refocus an input fluorescence image to different planes within the sample volume. Note that, even though *Deep-Z* was trained with uniform DPMs, in the testing phase one can also use spatially *non-uniform* entries as part of a DPM to refocus an input fluorescence image onto user-defined 3D surfaces, thanks to the convolutional nature of the network. This capability enables digital refocusing of the fluorescence image from a 3D surface onto another 3D surface, defined by the pixel mapping of the corresponding DPM.

Such a unique capability can be useful, among many applications, for simultaneous auto-focusing of different parts of a fluorescence image after the image capture, measurement or assessment of the aberrations introduced by the optical system (and/or the sample) as well as for correction of such aberrations by applying a desired non-uniform DPM. To exemplify this additional degree-of-freedom enabled by *Deep-Z*, **Figure 2.19** demonstrates the correction of the planar tilting and cylindrical curvature of two different samples, after the acquisition of a single 2D fluorescence image per object. **Figure 2.19(a)** illustrates the first measurement, where the plane of a fluorescent nano-bead sample was tilted by 1.5° with respect to the focal plane of the objective lens. As a result, the left and right sides of the acquired raw fluorescence image (**Figure 2.19(c)**) were blurred and the corresponding lateral FWHM values for these nano-beads became significantly wider, as reported in **Figure 2.19(e)**. By using a non-uniform DPM (see **Figure 2.19(b)**), which represents this sample tilt, *Deep-Z* can act on the blurred input image (**Figure 2.19(c)**) and accurately bring all the nano-beads into focus (**Figure 2.19(d)**), even though it was only trained using uniform DPMs. The lateral FWHM values calculated at the network output image became monodispersed, with a median of ~ 0.96 μm (**Figure 2.19(f)**), in comparison to a median of ~ 2.14 μm at the input image. Similarly, **Figure 2.19(g)** illustrates the second measurement, where the nano-beads were distributed on a *cylindrical* surface with a diameter of ~7.2 mm. As a result, the measured raw fluorescence image exhibited defocused regions as illustrated in **Figure 2.19(i)**, and the FWHM values of these nano-bead images were accordingly broadened (**Figure 2.19(k)**), corresponding to a median value of ~ 2.41 μm. On the other hand, using a non-uniform DPM that defines this cylindrical surface (**Figure 2.19(h)**), the aberration in **Figure 2.19(i)** was corrected using *Deep-Z* (**Figure 2.19(j)**), and similar to the

tilted sample case, the lateral FWHM values calculated at the network output image once again became monodispersed, as desired, with a median of ~ 0.91 μm (**Figure 2.19(l)**).

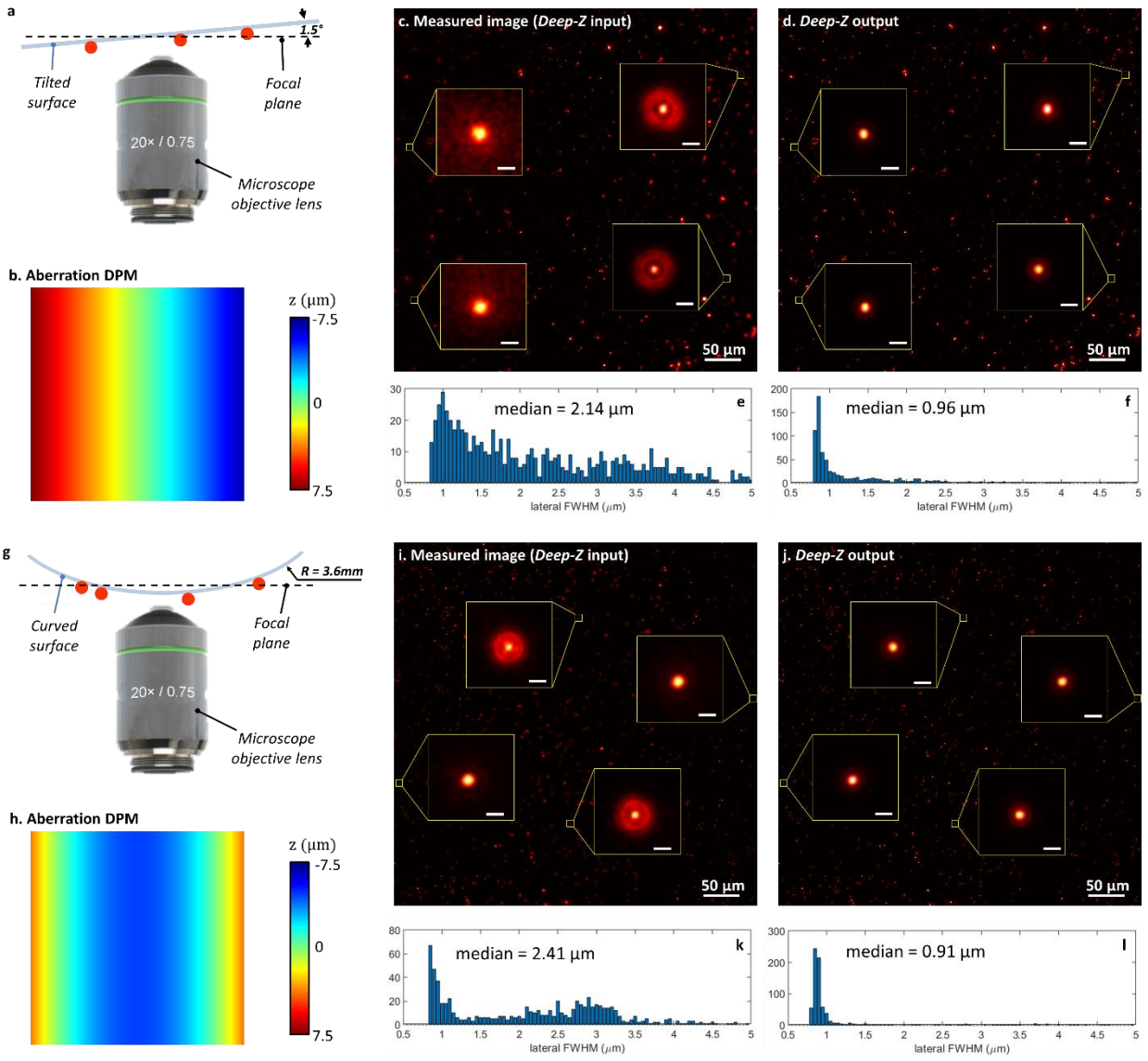

Figure 2.19 Non-uniform DPMs enable digital refocusing of a single fluorescence image onto user-defined 3D surfaces using *Deep-Z*. (a) Measurement of a tilted fluorescent sample (300 nm beads). (b) The corresponding DPM for this tilted plane. (c) Measured raw fluorescence image; the left and right parts are out-of-focus in different directions, due to the sample tilt. (d) The *Deep-Z* output rapidly brings all the regions into correct focus. (e,f) report the lateral FWHM values of the nano-beads shown in (c,d), respectively, clearly demonstrating that *Deep-Z* with the non-uniform DPM brought the out-of-focus particles into focus. (g) Measurement of a cylindrical surface with

fluorescent beads (300 nm beads). (h) The corresponding DPM for this curved surface. (i) Measured raw fluorescence image; the middle region and the edges are out-of-focus due to the curvature of the sample. (j) The *Deep-Z* output rapidly brings all the regions into correct focus. (k,l) report the lateral FWHM values of the nano-beads shown in (i,j), respectively, clearly demonstrating that *Deep-Z* with the non-uniform DPM brought the out-of-focus particles into focus. Scale bar in zoomed regions: 2 μm. Reprinted from Ref. [123].

Cross-modality digital refocusing of fluorescence images: *Deep-Z+*

Deep-Z framework enables digital refocusing of out-of-focus 3D features in a wide-field fluorescence microscope image to user-defined surfaces. The same concept can also be used to perform *cross-modality* digital refocusing of an input fluorescence image, where the generator network can be trained using pairs of input and label images captured by *two different* fluorescence imaging modalities, which is termed as *Deep-Z+*. After its training, the *Deep-Z+* network learns to digitally refocus a single input fluorescence image acquired by a fluorescence microscope to a user-defined target surface in 3D, but this time the output will match an image of the same sample captured by a *different* fluorescence imaging modality at the corresponding height/plane. To demonstrate this unique capability, I trained a *Deep-Z+* network using pairs of wide-field microscopy images (used as inputs) and confocal microscopy images at the corresponding planes (used as ground truth labels) to perform cross-modality digital refocusing (see the Methods for training details). **Figure 2.20** demonstrates the blind testing results for imaging microtubule structures of BPAEC using this *Deep-Z+* framework. As seen in **Figure 2.20**, the trained *Deep-Z+* network digitally refocused the input wide field fluorescence image onto different axial distances, while at the same time rejecting some of the defocused spatial features at the refocused planes, matching the confocal images of the corresponding planes, which serve as our ground truth. For instance, the microtubule structure at the lower left corner of a ROI in **Figure 2.20**, which was prominent at a refocusing distance of z = 0.34 μm, was

digitally rejected by *Deep-Z+* at a refocusing distance of z = -0.46 µm since it became out-of-focus at this axial distance, matching the corresponding image of the confocal microscope at the same depth. As demonstrated in **Figure 2.20**, *Deep-Z+* merges the sectioning capability of confocal microscopy with its image refocusing framework. **Figure 2.20** also reports *x-z* and *y-z* cross-sections of the *Deep-Z+* output images, where the axial distributions of the microtubule structures are significantly sharper in comparison to the axial scanning images of a wide-field fluorescence microscope, providing a very good match to the cross-sections obtained with a confocal microscope, matching the aim of its training.

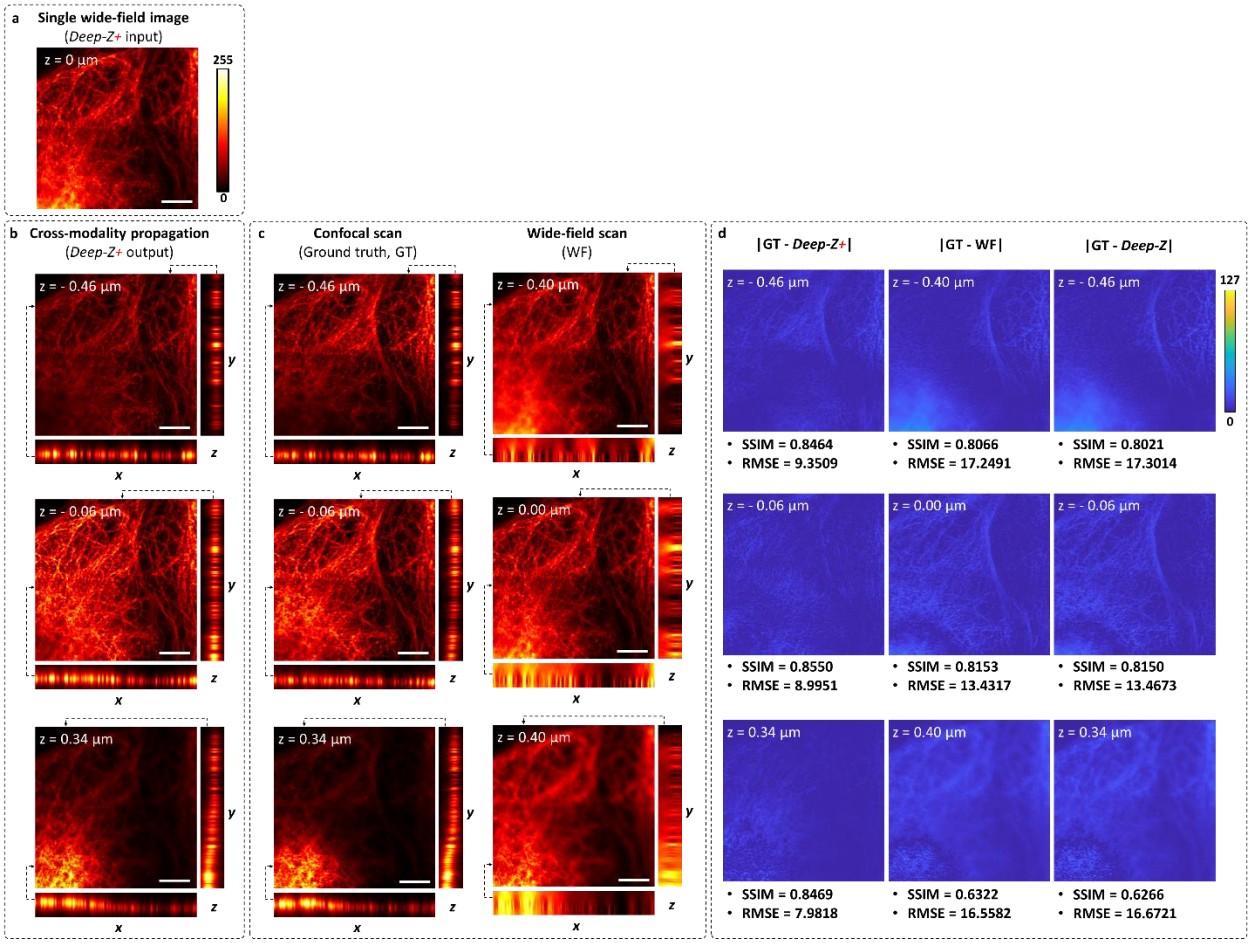

Figure 2.20 *Deep-Z+*: Cross-modality digital refocusing of fluorescence images. A single wide-field fluorescence image (63×/1.4NA objective lens) of BPAEC mgicrotubule structures (a) was digitally refocused using

Deep-Z+ to different planes in 3D (b), matching the images captured by a confocal microscope at the corresponding planes (c), retrieving volumetric information from a single input image and performing axial sectioning at the same time. Wide-field (WF) images are also shown in (c) for comparison. These scanning WF images report the closest heights to the corresponding confocal images, and have 60 nm axial offset since the two image stacks are discretely scanned and digitally aligned to each other. x-z and y-z cross-sections of the refocused images are also shown to demonstrate the match between *Deep-Z+* inference and the ground truth confocal microscope images of the same planes; the same cross-sections (x-z and y-z) are also shown for a wide-field scanning fluorescence microscope, reporting a significant axial blur in each case. Each cross-sectional zoomed-in image spans 1.6 μm in z-direction (with an axial step size of 0.2 μm), and the dotted arrows mark the locations, where the x-z and y-z cross-sections were taken. (d) The absolute difference images of the *Deep-Z+* output with respect to the corresponding confocal images are also provided, with SSIM and RMSE values, further quantifying the performance of *Deep-Z+*. For comparison, we also show the absolute difference images of the 'standard' *Deep-Z* output images as well as the scanning wide-field fluorescence microscope images with respect to the corresponding confocal images, both of which report increased error and weaker SSIM compared to |GT - *Deep-Z+*|. The quantitative match between |GT - WF| and |GT - *Deep-Z*| (see d) also suggests that the impact of 60 nm axial offset between the confocal and wide-field image stacks is negligible. Scale bar: 10 μm. Reprinted from Ref. [123].

2.5.3 Discussion

I developed a unique framework (*Deep-Z*), powered by deep neural networks, that enables 3D refocusing within a sample using a single 2D fluorescence image. This framework is non-iterative and does not require hyperparameter tuning following its training stage. In *Deep-Z*, the user can specify refocusing distances for each pixel in a DPM (following the axial range used in the training), and the fluorescence image can be digitally refocused to the corresponding surface through *Deep-Z*. The *Deep-Z* framework is also robust to changes in the density of the fluorescent objects within the sample volume, the exposure time of the input images, as well as the illumination intensity modulation (see Ref. [123] for detailed analyses). Because the distances are encoded in DPM and modeled as a convolutional channel, the network can be trained with

uniform DPMs, which still allows various non-uniform DPMs to be applied during the inference stage as reported in the Results section for e.g., correcting the sample drift, tilt, curvature or other optical aberrations, which brings additional degrees-of-freedom to the imaging system. Through this *Deep-Z* framework, I bring the snapshot *3D* refocusing capability of coherent imaging and holography to incoherent fluorescence microscopy, without any mechanical scanning, additional hardware components, or a trade-off of imaging resolution or speed. This not only significantly boosts the imaging speed, but also reduces the negative effects of photobleaching and phototoxicity on the sample, which is a crucial issue for fluorescence microscopy experiments.

Practically, the retrievable axial range in this method depends on the SNR of the recorded image, i.e., if the depth information carried by the PSF falls below the noise floor, accurate inference will become a challenging task. In Results section, I demonstrated an enhancement of ~20× in the DOF of a wide-field fluorescence image using *Deep-Z*. This axial refocusing range is in fact not an absolute limit but rather a practical choice for my training data, and it may be further improved through hardware modifications to the optical set-up by e.g., engineering the PSF in the axial direction [211,222,235–237]. In addition to requiring extra hardware and sensitive alignment/calibration, such approaches would also require brighter fluorophores, to compensate for photon losses due to the insertion of additional optical components in the detection path.

2.5.4 Methods

Sample preparation

Structural imaging of *C. elegans* neurons was carried out in strain AML18. AML18 carries the genotype wtfIs3 [rab-3p::NLS::GFP + rab-3p::NLS::tagRFP] and expresses GFP and tagRFP in the nuclei of all the neurons [232]. For functional imaging, I used the strain AML32, carrying

wtfIs5 [rab-3p::NLS::GCaMP6s + rab-3p::NLS::tagRFP]. The strains were acquired from the Caenorhabditis Genetics Center (CGC). Worms were cultured on Nematode Growth Media (NGM) seeded with OP50 bacteria using standard conditions [238]. For imaging, worms were washed off the plates with M9, and anaesthetized with 3 mM levamisole [239]. Anaesthetized worms were then mounted on slides seeded with 3% Agarose. To image moving worms, the levamisole was omitted.

Fluorescence image acquisition

The fluorescence images of nano-beads, *C. elegans* structure and BPAEC samples were captured by an inverted scanning microscope (IX83, Olympus Life Science) using a 20×/0.75NA objective lens. A 130W fluorescence light source was used at 100% output power. Two bandpass optical filter sets were used: Texas Red and FITC. For all the samples, the scanning microscope had a motorized stage that moved the samples to different FOVs and performed image-contrast-based auto-focus at each location. The motorized stage was controlled using MetaMorph® microscope automation software (Molecular Devices, LLC). At each location, the control software autofocused the sample based on the standard deviation of the image, and a z-stack was taken from -20 μm to 20 μm with a step size of 0.5 μm. The image stack was captured by a monochrome scientific CMOS camera (ORCA-flash4.0 v2, Hamamatsu Photonics K.K), and saved in non-compressed tiff format, with 81 planes and 2048 × 2048 pixels in each plane.

The images of *C. elegans* neuron activities were captured by another scanning wide-field fluorescence microscope (TCS SP8, Leica Microsystems) using a 20×/0.8NA objective lens and a 40×/1.3NA objective lens. Two bandpass optical filter sets were used: Texas Red and FITC. The images were captured by a monochrome scientific CMOS camera (Leica-DFC9000GTC-VSC08298). For capturing image stacks of anesthetized worms, the motorized stage controlled

by a control software (LAS X, Leica Microsystems) moved the sample slide to different FOVs. At each FOV, the control software took a z-stack from -20 μm to 20 μm with a step size of 0.5 μm for the 20×/0.8NA objective lens images, and with a step size of 0.27 μm for the 40×/1.3NA objective lens images, with respect to a middle plane (z=0 μm). Two images were taken at each z-plane, for Texas Red channel and FITC channel respectively. For capturing 2D videos of dynamic worms, the control software took a time-lapsed video that also time-multiplexed the Texas Red and FITC channels at the maximum speed of the system.

The BPAEC wide-field and confocal fluorescence images for *Deep-Z+* were captured by another inverted scanning microscope (TCS SP5, Leica Microsystems). The images were acquired using a 63×/1.4 NA objective lens and FITC filter set was used. The wide-field images were recorded by a CCD with 1380×1040 pixels and 12-bit dynamic range, whereas the confocal images were recorded by a photo-multiplier tube (PMT) with 8-bit dynamic range (1024×1024 pixels). The scanning microscope had a motorized stage that moved the sample to different FOVs and depths. At each location, a stack of 12 images with 0.2 μm axial spacing was recorded.

Image pre-processing and training data preparation

Each captured image stack was first axially aligned using an ImageJ plugin named "StackReg" [202], which corrects the rigid shift and rotation within the stack. Then an extended depth of field (EDF) image was generated using another ImageJ plugin named "Extended Depth of Field" [240]. This EDF image was used as a reference image to normalize the whole image stack, following three steps: (1) Triangular threshold[241] was used on the image to separate the background and foreground pixels; (2) the mean intensity of the background pixels of the EDF image was determined to be the background noise and subtracted; (3) the EDF image intensity was scaled to 0-1, where the scale factor was determined such that 1% of the foreground pixels above the

background were greater than one (i.e., saturated); and (4) each image in the stack was subtracted by this background level and normalized by this intensity scaling factor. For testing data without an image stack, steps (1) – (3) were applied on the input image instead of the EDF image.

To prepare the training and validation datasets, on each FOV, a geodesic dilation [104] with fixed thresholds was applied on fluorescence EDF images to generate a mask that represents the regions containing the sample fluorescence signal above the background. Then, a customized greedy algorithm was used to determine a minimal set of regions with 256 × 256 pixels that covered this mask, with ~5% area overlaps between these training regions. The lateral locations of these regions were used to crop images on each height of the image stack, where the middle plane for each region was set to be the one with the highest standard deviation. Then 20 planes above and 20 planes below this middle plane were set to be the range of the stack, and an input image plane was generated from each one of these 41 planes. Depending on the size of the data set, around 5-10 out of these 41 planes were randomly selected as the corresponding target plane, forming around 150 to 300 image pairs. For each one of these image pairs, the refocusing distance was determined based on the location of the plane (i.e. 0.5 μm times the difference from the input plane to the target plane). By repeating this number, a uniform DPM was generated and appended to the input fluorescence image. The final dataset typically contained ~ 100,000 image pairs. This was randomly divided into a training dataset and a validation dataset, which took 85% and 15% of the data respectively. During the training process, each data point was further augmented five times by flipping or rotating the images by a random multiple of 90°. The validation dataset was not augmented. The testing dataset was cropped from separate measurements with sample FOVs that do not overlap with the FOVs of the training and validation data sets.

Network architecture and training

The *Deep-Z* network is formed by a least square generative adversarial network (LS-GAN) framework [242], and it is composed of two parts: a generator and a discriminator, as shown in **Figure 2.6**. The generator is a convolutional neural network (CNN) inspired by the U-Net [196], as detailed in section 2.2.2. The discriminator is a convolutional neural network that consists of six consecutive convolutional blocks. The *Deep-Z* network learns to use the information given by the appended DPM to digitally refocus the input image to a user-defined plane. In the training phase, the input data of the generator have the dimensions of 256 × 256 × 2, where the first channel is the fluorescence image, and the second channel is the user-defined DPM. The target data of generator have the dimensions of 256 × 256, which represent the corresponding fluorescence image at a surface specified by the DPM. The input data of the discriminator $D(.)$ have the dimensions of 256 × 256, which can be either the generator output or the corresponding target. During the training phase, the network iteratively minimizes the generator loss L_G and discriminator loss L_D, defined as in equations (1.11) and (1.13). Adam optimizer was used to minimize both L_G and L_D, with a learning rate of 10^{-4} and 3×10^{-5} for L_G and L_D respectively. In each iteration, six updates of the generator loss and three updates of the discriminator loss were performed. The validation set was tested every 50 iterations, and the best network (to be blindly tested) was chosen to be the one with the smallest MAE loss on the validation set. The network was implemented using TensorFlow [110], performed on a PC with Intel Core i7-8700K six-core 3.7GHz CPU and 32GB RAM, using an Nvidia GeForce 1080Ti GPU. On average, the training takes ~ 70 hours for ~ 400,000 iterations (equivalent to ~ 50 epochs). After the training, the network inference time was ~ 0.2 s for an image with 512 × 512 pixels and ~ 1s for an image with 1536 × 1536 pixels on the same PC.

Tracking and quantification of *C. elegans* neuron activity

To localize individual neurons, the red channel stacks (Texas Red, neuron nuclei) were projected by median-intensity through the time sequence. Local maxima in this projected median intensity stack marked the centroid of each neuron and the voxels of each neuron was segmented from these centroids by watershed segmentation[104], which generated a 3D spatial voxel mask for each neuron. A total of 155 neurons were isolated. Then, the average of the 100 brightest voxels in the green channel (FITC, neuron activity) inside each neuron spatial mask was calculated as the calcium activity intensity $F_i(t)$, for each time frame t and each neuron $i = 1,2,\dots,155$. The differential activity was then calculated, $\Delta F(t) = F(t) - F_0$, for each neuron, where F_0 is the time average of $F(t)$.

By thresholding on the standard deviation of each $\Delta F(t)$, I selected the 70 most active cells and performed further clustering on them based on their calcium activity pattern similarity using a spectral clustering algorithm [243,244]. The calcium activity pattern similarity was defined as

$$S_{ij} = \exp\left(-\frac{\left\| \frac{\Delta F_i(t)}{F_{i0}} - \frac{\Delta F_j(t)}{F_{j0}} \right\|^2}{\sigma^2} \right) \tag{2.14}$$

for neurons i and j, which results in a similarity matrix S. $\sigma = 1.5$ is the standard deviation of this Gaussian similarity function, which controls the width of the neighbors in the similarity graph. The spectral clustering solves an *eigen-value problem* on the graph Laplacian L generated from the similarity matrix S, defined as the difference of weight matrix W and degree matrix D,

$$L = D - W \tag{2.15}$$

where

$$W_{ij} = \begin{cases} S_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{2.16}$$

$$D_{ij} = \begin{cases} \sum_{j} W_{ij} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.17}$$

The number of clusters was chosen using eigen-gap heuristics[244], which was the index of the largest general eigenvalue (by solving general eigen value problem $Lv = \lambda Dv$) before the eigen-gap, where the eigenvalues jump up significantly, which was determined to be k=3. Then the corresponding first k=3 eigen-vectors were combined as a matrix, whose rows were clustered using standard k-means clustering[244], which resulted in the three clusters of the calcium activity patterns shown in **Figure 2.18**.

Cross-modality alignment of wide-field and confocal fluorescence images

Individual FOVs of each stack of the wide-field/confocal pair were stitched together using "Image Stitching" plugin of ImageJ [245]. The stitched wide-field and confocal EDF images were then co-registered using a feature-based registration with projective transformation performed in Matlab (MathWorks, Inc) [203]. Then the stitched confocal EDF images as well as the stitched stacks were warped using this estimated transformation to match their wide-field counterparts (**Figure 2.21(a)**). The non-overlapping regions of the wide-field and warped confocal images were subsequently deleted. Then the above-described greedy algorithm was used to crop non-empty regions of 256 × 256 pixels from the remaining stitched wide-field images and their corresponding warped confocal images. The same feature-based registration was applied on each pair of cropped regions for fine alignment. This step provides good correspondence between the wide field image and the corresponding confocal image in the lateral directions (**Figure 2.21(b)**).

Although the axial scanning step size was fixed to be 0.2 μm, the reference zero-point in the axial direction for the wide-field and the confocal stacks needed to be matched. To determine this reference zero-point in the axial direction, the images at each depth were compared with the EDF image of the same region using structural similarity index (SSIM) [181], providing a focus curve (**Figure 2.21(c)**). A second order polynomial fit was performed on four points in this focus curve with highest SSIM values, and the reference zero-point was determined to be the peak of the fit (**Figure 2.21(c)**). The heights of wide-field and confocal stacks were then centered by their corresponding reference zero-points in the axial direction. For each wide-field image used as input, four confocal images were randomly selected from the stack as the target, and their DPMs were calculated based on the axial difference of the centered height values of the confocal and the corresponding wide-field images.

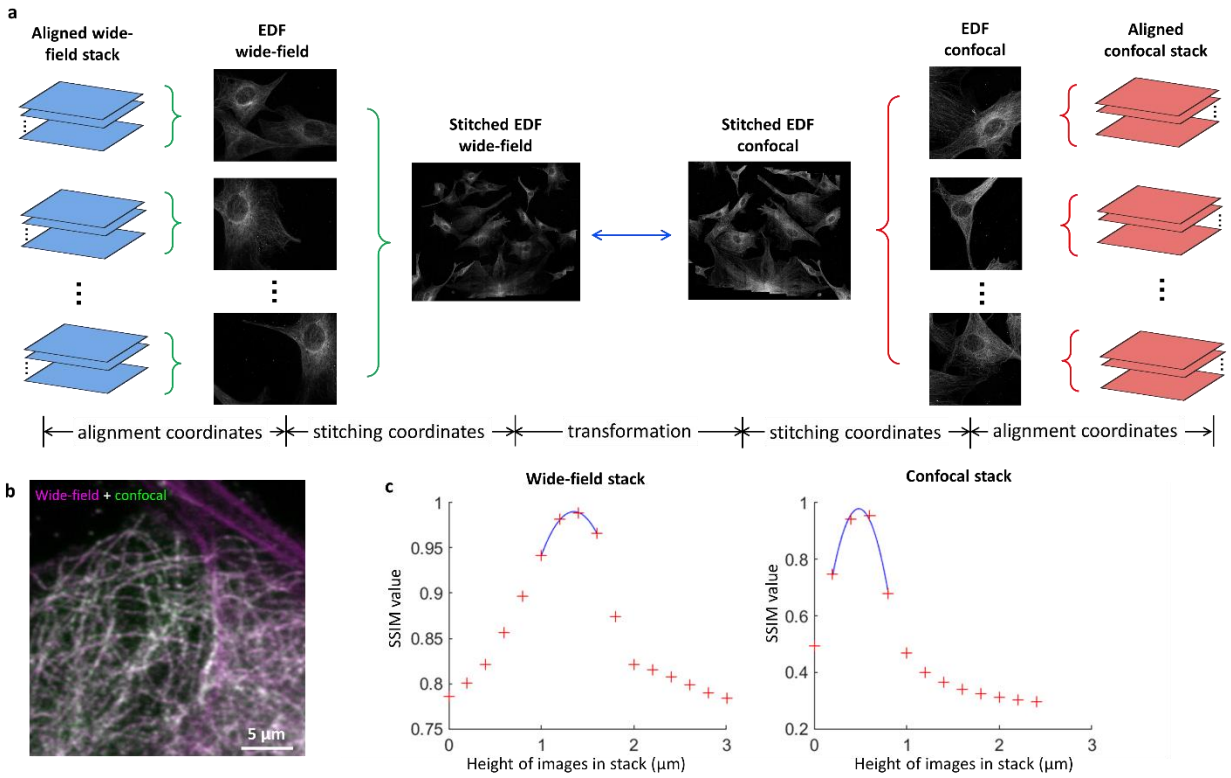

Figure 2.21 Registration of a wide-field fluorescence z-stack against a confocal z-stack. (a) Registration in the lateral direction. Both the wide-field and the confocal z-stacks were first self-aligned and extended depth of field (EDF) images were calculated for each stack. The EDF images were stitched spatially and the stitched EDF images from wide-field were aligned to those of confocal microscopy images. The spatial transformations, from stitching to the EDF alignment, were used as consecutive transformations to associate the wide-field stack to the confocal stack. (b) Non-empty wide-field ROIs of 256×256 pixels and the corresponding confocal ROIs were cropped from the EDF image, which were further aligned. The image here shows an example overlay of the registered image pair, with wide-field image in magenta and the corresponding confocal image in green. (c) To align the wide-field and confocal stacks in the axial direction, focus curves in the wide-field stack and the confocal stack were calculated and compared based on the corresponding SSIM values.

Chapter 3 Deep learning enabled air quality monitoring

One of the major sources of air pollution is created by particulate matter (PM), which refers to the suspended biological or non-biological particles in air. PM in general can penetrate human respiration system and even blood and cause allergies, various diseases and even cancer. Currently, a majority of PM monitoring processes involve capturing samples in the field and analyzing the collected PMs in a remote laboratory, where a micro-biologist or microscopist examines the sample, typically using a microscope. These manual lab inspection-based approaches are very slow, time-consuming and expensive, thus there is an unmet need for mobile and cost-effective technologies that can rapidly and accurately measure the concentration of such bio-aerosols in the field, away from a lab.

Using computational lens-less digital holographic microscopy, powered by machine learning and deep learning, I created a hand-held, high-throughput, and intelligent PM monitor, called c-Air (pronounced as "see air", where the letter "c" stands for "computation"). In section 3.1, I will describe accurate PM detection and sizing using c-Air device. In section 3.2, I will describe automatic and label-free bio-aerosol sensing using c-Air device.

Part of this chapter has been previously published in

- Y. Wu et al. "Air Quality Monitoring Using Mobile Microscopy and Machine Learning." Light: Science & Applications 6, no. 9 (September 2017): e17046;
- Y. Wu et al. "Label-Free Bioaerosol Sensing Using Mobile Microscopy and Deep Learning." ACS Photonics 5, no. 11 (November 21, 2018): 4617–27.

3.1 Particulate matter monitoring

Rapid, accurate, and high-throughput sizing and quantification of particulate matter (PM) in air is crucial for monitoring and improving air quality. Here, I demonstrate a field-portable cost-effective platform for high-throughput quantification of particulate matter using computational lens-less microscopy and machine-learning. This platform, termed c-Air, is also integrated with a smartphone application for device control and display of results. This mobile device rapidly screens 6.5 L of air in 30 s and generates microscopic images of the aerosols in air. It provides statistics of the particle size and density distribution with a sizing accuracy of ~93%. I tested this mobile platform by measuring the air quality at different indoor and outdoor environments and measurement times, and compared our results to those of an Environmental Protection Agency–approved device based on beta-attenuation monitoring, which showed strong correlation to c-Air measurements. Furthermore, I used c-Air to map the air quality around Los Angeles International Airport (LAX) over 24 hours to confirm that the impact of LAX on increased PM concentration was present even at >7 km away from the airport, especially along the direction of landing flights. With its machine-learning based computational microscopy interface, c-Air can provide a cost-effective mobile solution for highly accurate and distributed sensing of air quality.

3.1.1 Introduction

Particulate matter (PM) is a mixture of solid and liquid particles in air and forms a significant form of air pollution. PM sources include, for example, direct emission from a source, such as a construction site, smokestack, or fire, or a result of complex chemical reactions emitted from power plants, industrial production, and automobiles.[246] PM with a general diameter of 10 μm and smaller, which is termed PM10, can cause serious health problems because it can become lodged deep in the lungs and even enter the bloodstream. A smaller PM size category, PM2.5,

which represents particles with a diameter of 2.5 μm or smaller, has been declared a cause of cancer by the WHO.[247] Furthermore, PM is a major environmental issue on account of reduced visibility (haze). Monitoring PM air quality as a function of space and time is critical for understanding the effects of industrial activities, studying atmospheric models, and providing regulatory and advisory guidelines for transportation, residents, and industries.

Currently, PM monitoring is performed at designated air sampling stations, which are regulated by the US Environmental Protection Agency (EPA) and similar agencies in different countries. Many of these advanced automatic platforms use either beta-attenuation monitoring (BAM)[248] or a tapered element oscillating microbalance (TEOM)[248] instrument. BAM instruments sample aerosols on a rotating filter. Using a beta-particle source, they measure the beta-particle attenuation induced by the accumulated aerosols on the filter. TEOM-based instruments, on the other hand, capture aerosols in a filter cartridge, which contains a glass tube tip that vibrates at varying frequencies according to the mass of the captured aerosols. These devices provide accurate PM measurements at high throughputs. However, they are cumbersome and heavy (~30 kg), relatively expensive (~$50,000–100,000), and require specialized personnel or technicians for regular system maintenance, e.g., every few weeks. Owing to these limitations, only approximately 10,000 of these air sampling stations exist worldwide.

In addition to these high-end PM measurement instruments, several commercially available portable particle counters are available at a lower cost of approximately $2,000[249,250] and in some cases much higher, ~$7,000-8,000.[251] These commercially available optical particle counters resemble a flow-cytometer, which drives the sampled air through a small channel. A laser beam focused on the nozzle of this channel is scattered by each particle that passes through the channel. The scattering intensity is then used to infer the particle size. Because of its serial read-out nature,

the sampling rate of this approach is limited to < 2-3 L/min and in some sub-micron particle detection schemes <0.8 L/min.[251] Furthermore, accurate measurement of either very high or low concentrations of particles is challenging for these devices, which limits the dynamic range of the PM measurement. Aside from these limitations, the scattering cross-section, which comprises the quantity actually measured by this device type, heavily depends on the 3D morphology and refractive properties of the particles. This can cause severe errors in the conversion of the measured scattered light intensities into actual particle sizes. Finally, none of these designs offers a direct measure, i.e., a microscopic image of the captured particles, which is another limitation because further analysis of a target particle of interest after its detection cannot be performed.

On account of these limitations, many air sampling activities continue to use microscopic inspection and counting. Basically, air samples are manually obtained in the field using a portable sampler that employs various processes, such as cyclonic collection, impingement, impaction, thermophoresis, or filtering.[252–255] The sample is then sent to a central laboratory, where it is post-processed and manually inspected under a microscope by an expert. This type of microscopic analysis provides the major advantage of more accurate particle sizing, while enabling the expert reader to recognize the particle shape and type. These capabilities yield additional benefits in more complicated analyses of air pollution, such as identification of a specific aerosol type. In this method, however, the sampling and inspection processes are separated; i.e., *the sampling is performed in the field, whereas the sample analysis is conducted in a remote professional lab*. This significantly delays the reporting of the results. Furthermore, the inspection is manually performed by a trained expert, which considerably increases the overall cost of each air-quality measurement. Furthermore, this conventional microscope-based

screening of aerosols cannot be conducted in the field, because these benchtop microscopes are cumbersome, heavy, expensive, and require specialized skills to operate.

As a transformative solution to the above outlined limitations of existing air quality measurement techniques, I developed a hand-held and cost-effective platform for automated sizing and high-throughput quantification of PM using computational lens-less microscopy and machine learning. As an alternative to conventional lens-based microscopy techniques, in a computational lens-less microscopy approach, the sample is directly positioned on top of an image sensor chip with no optical components between them.[15,16] Such an on-chip microscope can computationally reconstruct images of transmissive samples over a very large field-of-view of >20 mm^2. Based on this computational on-chip imaging concept and a unique machine learning enabled particle analysis method, I demonstrate the design of c-Air -- a lightweight (~590 g), hand-held and cost-effective air-quality monitoring system. This mobile system utilizes a micro-pump, an impaction-based air-sampler, and a lens-less holographic on-chip microscope that is integrated with a custom-written machine learning algorithm for remote data processing and particle analysis. The c-Air platform (**Figure 3.1** and **Figure 3.2**) operates with a smartphone application (app) for device control and data display. It can rapidly screen 6.5 L of air volume in 30 s, generating microscopic phase and amplitude images of the captured aerosols, while also automatically providing the PM sizing distribution with a sizing accuracy of ~93%.

By measuring the air quality using c-Air over several hours at an EPA-approved air-quality monitoring station, I confirmed that the c-Air PM measurements closely matched those of an EPA-approved BAM device. I further tested c-Air prototypes in various indoor and outdoor locations in California. In some of these experiments, a significant increase in ambient PM caused by the so-called Sand Fire near Santa Clarita, California on July 23, 2016 was detected.

Finally, I conducted a 24-h spatio-temporal mapping of air pollution near Los Angeles International Airport (LAX) using the c-Air prototypes. The results revealed the occurrence of a temporal modulation of PM that correlates with the total number of flights at LAX. This modulation was present even at a distance of > 7 km from LAX along the direction of landing flights. Because c-Air is based on computational microscopic imaging and machine learning, I believe that the c-Air platform and its unique capabilities are broadly applicable to numerous air-quality-related applications, and it can provide cost-effective, compact, and mobile solutions for spatio-temporal mapping of both indoor and outdoor air quality.

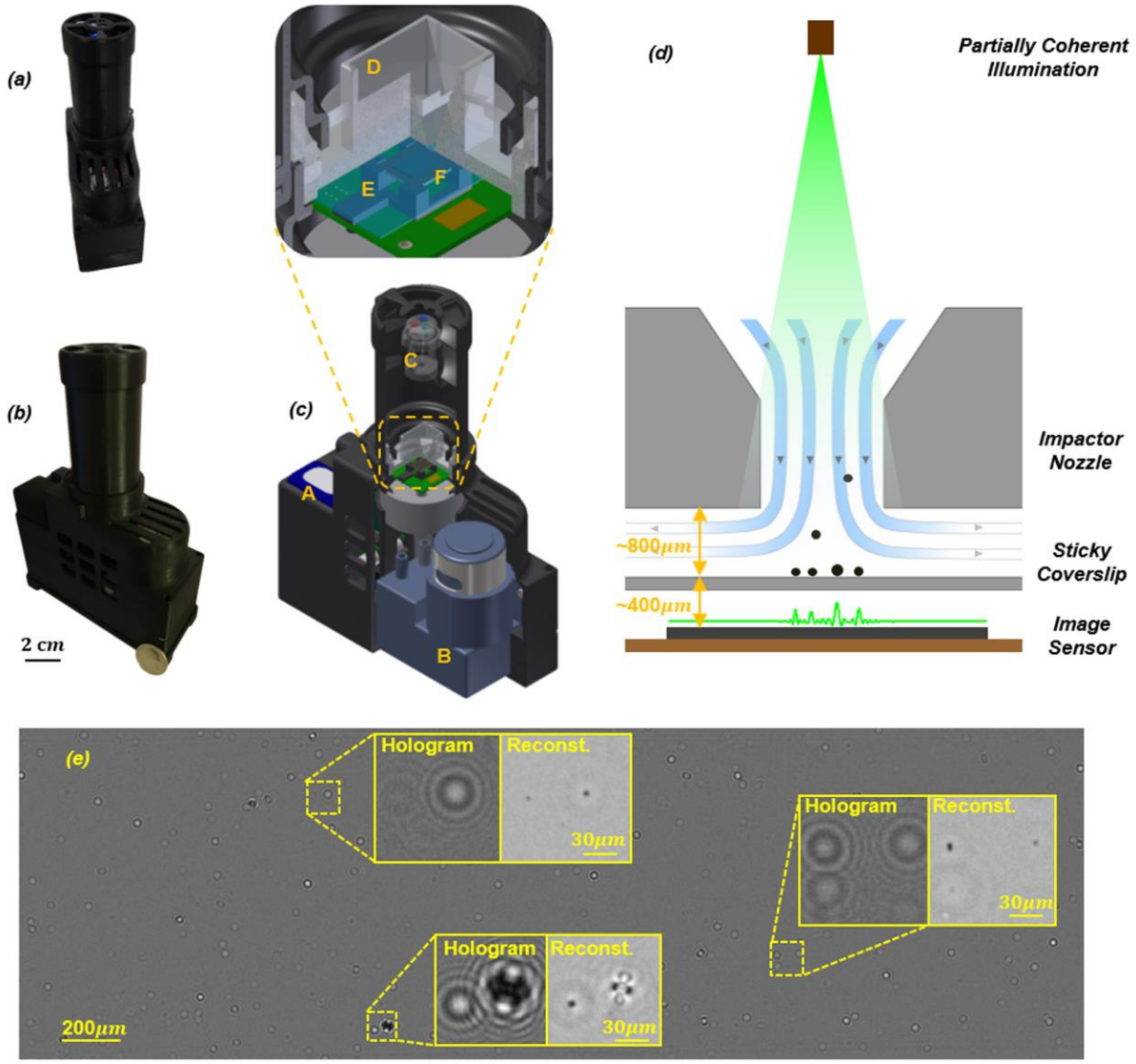

Figure 3.1 c-Air platform. (a)–(b) Photos of the device from different perspectives. A quarter is placed beside the device in (b) for scale. (c) 3D computer-aided-design (CAD)-drawing overview of the device, including (A) rechargeable battery, (B) vacuum pump (13 L/min), (C) illumination module with fiber-coupled LEDs of red (624 nm), green (527 nm), and blue (470 nm), (D) impaction-based air sampler with (E) a sticky coverslip on top of (F) the image sensor. (d) Schematic drawing of impaction-based air sampler on a chip. (e) Whole field-of-view differential hologram image of an aerosol sample during sampling, and zoomed-in regions of its reconstruction. The device is powered by a rechargeable battery (A), and controlled by a microcontroller (Raspberry Pi A+). The assembly and packaging are 3D-printed. Reprinted from Ref. [23]

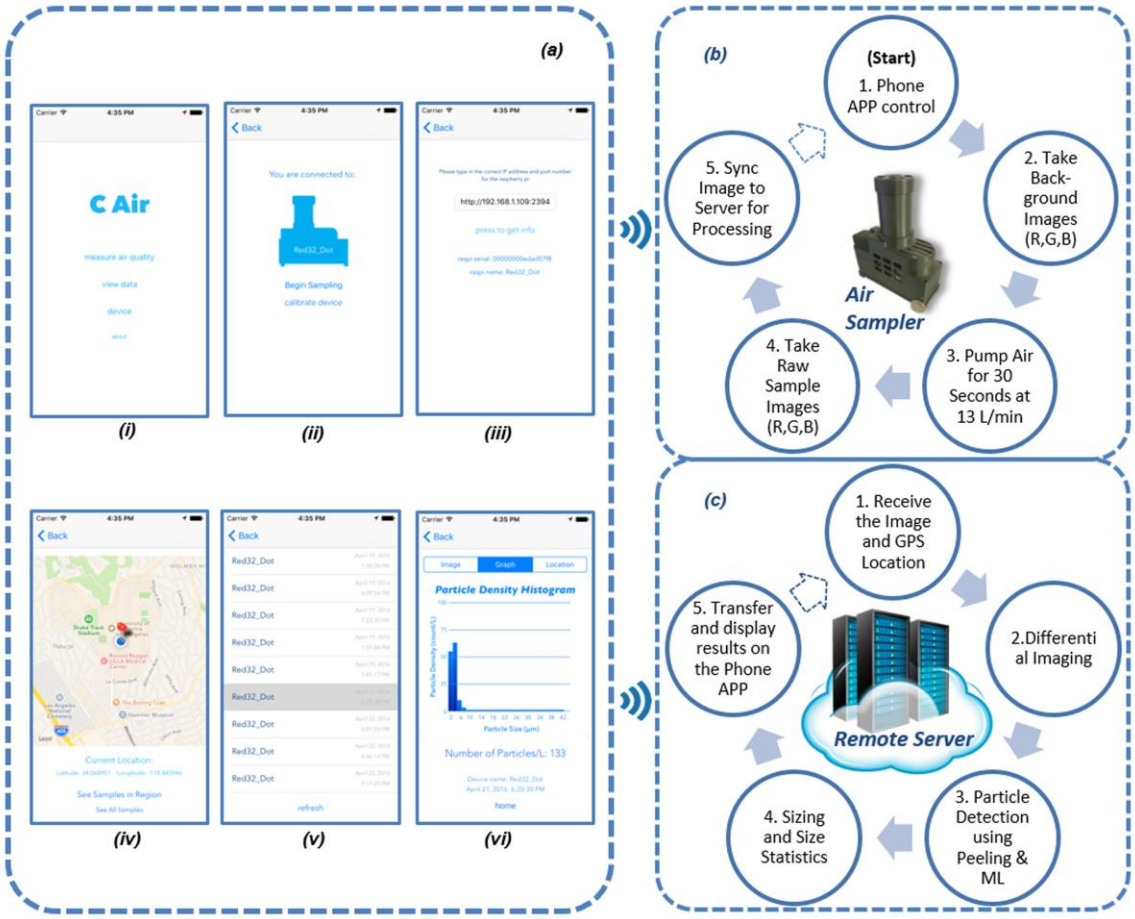

Figure 3.2 c-Air work flow and app interface. (a) iOS-based c-Air app interface: (i) "Welcome" screen of the app with different options. (ii) "Take Measurement" screen with a device-logo-shaped sampling button. (iii) Changing the device connection. The user can change the device to be connected by typing the IP address of the device. (iv) "Map View" of history samples. The air samples can be viewed by touching the pinpoint. (v) "List View" of history samples. Each entry is a sample that shows the device name and capture time. (vi) View of one sample result. The "graph" option shows a histogram of the particle sizing. (b) Workflow on the c-Air device. (c) Workflow on the server to support the processing of air samples. After the sample image and GPS location are sent to the server, the server processes the images, and the result is sent to the smartphone app, where it is rendered and displayed. (GPS: global position system; ML: machine learning). Reprinted from Ref. [23]

3.1.2 Results and discussion

Particle sizing accuracy using machine learning

I used a machine-learning algorithm trained on size-calibrated particles to obtain a mapping from the detected spatial characteristics of the particles to their diameter (see Methods in sub-section 3.1.4 for details). **Figure 3.3** depicts how well the predicted particle diameter, D_{pred}, based on this machine learning algorithm, agrees with the ground-truth measured diameter, D_{ref}. The sizing errors for training and testing sets are defined in equations (3.7) and (3.8), respectively. The dotted black line in **Figure 3.3** represents the reference for $D_{ref} = D_{pred}$. As shown in **Figure 3.3**, using machine learning, c-Air achieved an average sizing accuracy of approximately 93% for both the training and blind testing sets.

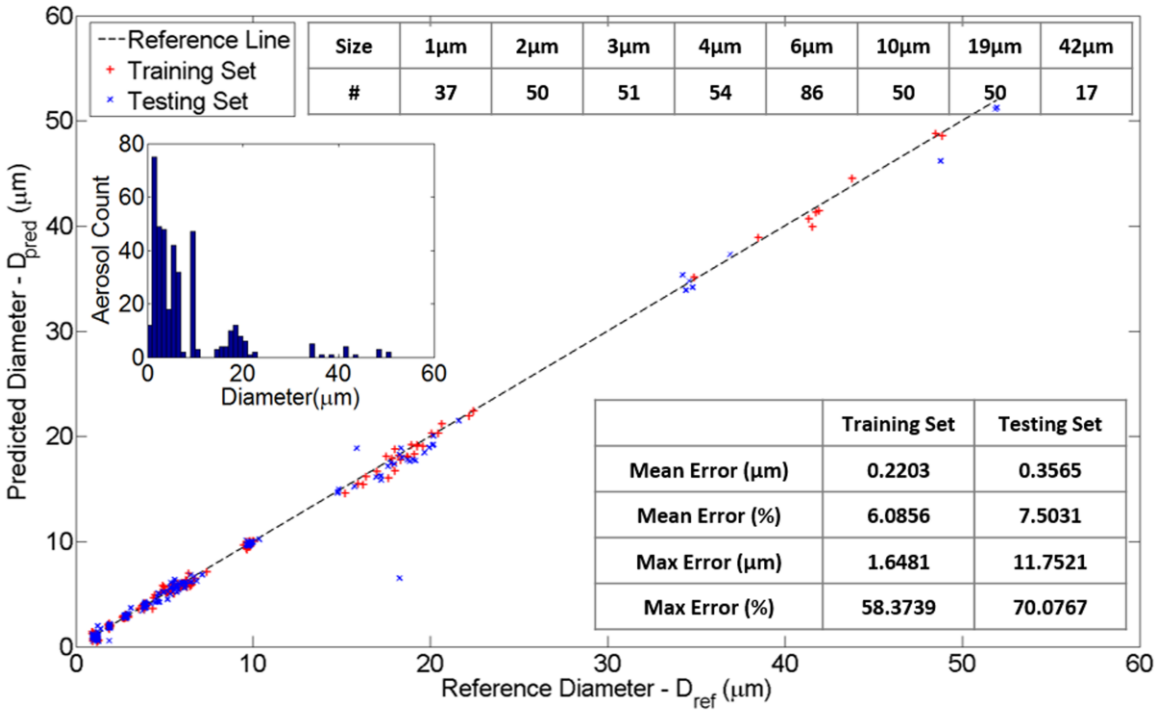

Figure 3.3 Machine learning based particle detection and sizing with high accuracy using c-Air. The designated bead sizes are shown in the uppermost table. The microscope-calibrated size distribution is plotted as the histogram within the large figure. The large figure in the background shows the machine-learning mapped size

(D_{pred}) using c-Air. It is compared to the microscope-measured size (D_{ref}) for both training and testing sets. The middle dashed line represents $D_{pred} = D_{ref}$. The sizing error, which is defined by equations (6) to (7), is shown in the lower-right table in both μm and the relative percentage. A ~93% accuracy using machine-learning-based sizing is demonstrated. Reprinted from Ref. [23]

Particle size distribution measurements and repeatability of the c-Air platform

I employed two c-Air devices (which were designed to be identical) and conducted repeated measurements at four locations: (1) the class-100 clean room of California NanoSystems Institute (CNSI) on 6/21/2016; (2) the class-1000 clean room at CNSI on 5/23/2016; (3) the indoor environment in the Engineering IV building at the University of California, Los Angeles (UCLA) campus on 5/25/2016; and (4) the outdoor environment at the second floor patio of the Engineering IV building on 05/23/2016. At each location, we obtained seven samples for each c-Air device with a sampling period of 30 s between the two successive measurements. These sample c-Air images were processed as described in the Methods section, and the particle size distributions for each location were analyzed and compared.

Figure 3.4 (a) to (c) show a box-whisker plot of the data distribution for total suspended particulate (TSP), PM10, and PM2.5 at these four locations. The points, marked by an "x" symbol, were excluded as outliers from the box plot with a whisker length of 1.5 (99.3% confidence). It is interesting to note that c-Air measured the TSP density at $\sim 7\ \mathrm{counts/L}$ for the class-100 clean room and $25\ \mathrm{counts/L}$ for the class-1000 clean room at CNSI, which are both comparable to the ISO 14644-1 clean room standards[256], i.e., $3.5\ \mathrm{counts/L}$ for the class-100 clean room and $36\ \mathrm{counts/L}$ for the class-1000 clean room for particles $\geq 0.5\ \mu\mathrm{m}$.

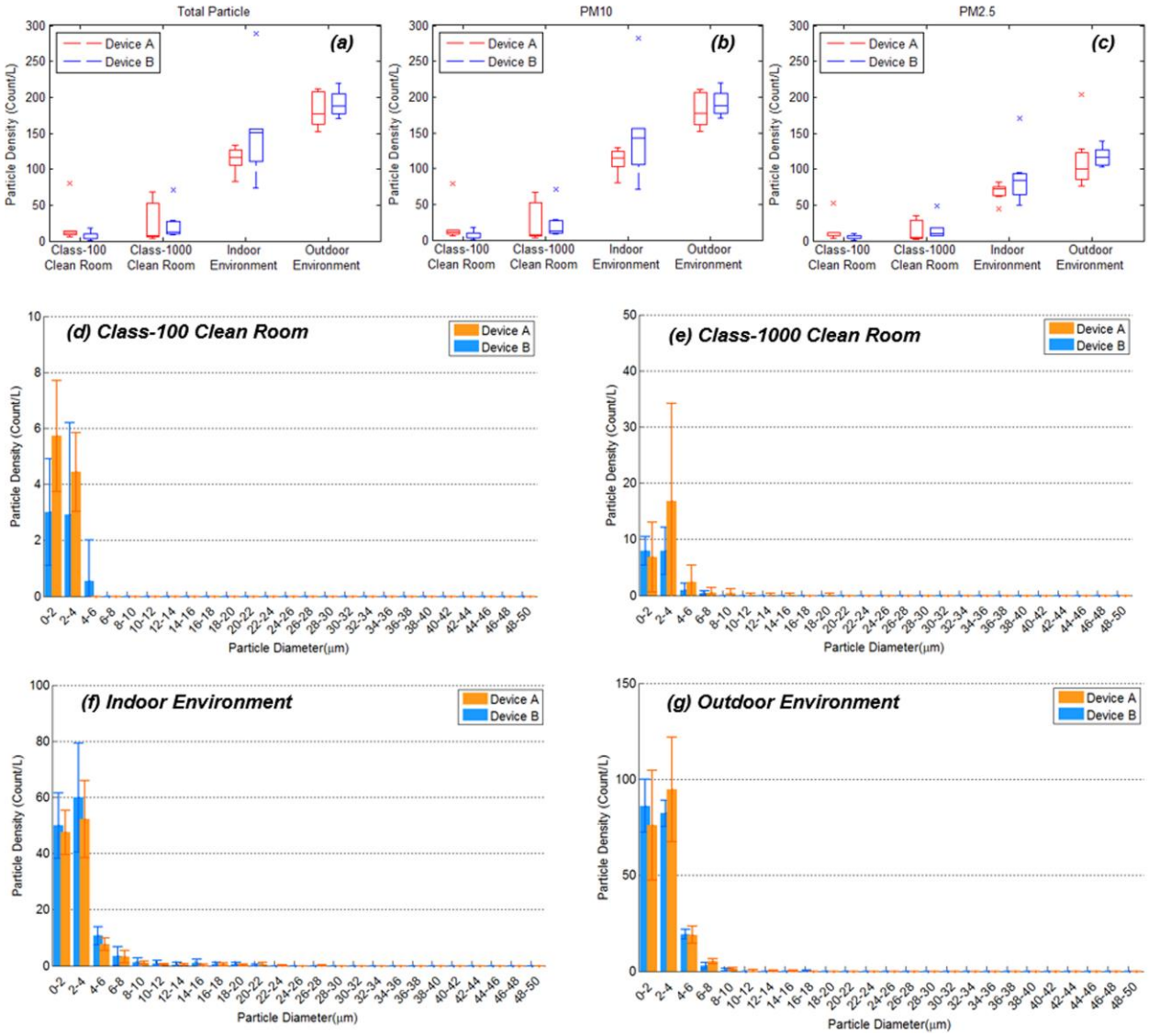

Figure 3.4 c-Air repeatability tests at different locations. (a)–(c) Box-plot of the repeatability test results using two c-Air devices (A and B) at the (1) class-100 clean room of CNSI on 06/21/2016; (2) class-1000 clean room at CNSI on 05/23/2016; (3) indoor environment of the UCLA Engineering IV building on 05/25/2016; and (4) outdoor environment at the second floor patio of the UCLA Engineering IV building on 05/23/2016. The box-plot was generated using the box-whisker plot method with a whisker length of 1.5 (99.3% confidence) to exclude outliers, which are marked by the symbol "x." (d)–(g) Particle size and density distribution histogram comparison at each location: (d) Class-100 clean room; (e) class-1000 clean room; (f) indoor environment; and (g) outdoor environment. Seven samples for each c-Air device with a sampling period of 30 s were obtained at each location. Reprinted from Ref. [23]

The measurements of TSP, PM10, and PM2.5 densities from the same data set were additionally used to elucidate two aspects of the repeatability of the c-Air platform, i.e., the intra-device and inter-device repeatability. The intra-device repeatability is defined as the extent to which the measurement result varies from sample to sample using the same c-Air device to measure the air quality in the same location (assuming that the air quality does not change from measurement to measurement with a small time lag in between). The strong intra-device repeatability of c-Air is evident in the "max–min" perspective in the box plot of **Figure 3.4**.

The inter-device repeatability is defined as the extent to which the results vary from each other using two c-Air devices that are designed as identical to measure the air quality in the same location. This can be qualitatively viewed by comparing the measurement result of device A and device B in **Figure 3.4**. To further quantify the inter-device repeatability, I performed a μ-test (i.e., Mann–Whitney μ-test or Wilcoxon rank sum test) on the 2×4 sets of measurement data from devices A and B at four different locations. In the μ-test, I aimed to verify the null hypothesis ($H = 0$) for two sets of samples, X and Y:

$$\mathrm{H} = 0 : \mathrm{P}(\mathrm{X} > \mathrm{Y}) = \mathrm{P}(\mathrm{Y} > \mathrm{X}) = \frac{1}{2} \tag{3.1}$$

that is, to test if the medians of the two samples are statistically the same. Compared to other tests for repeatability, e.g., the student t-test, the μ-test requires fewer assumptions and is more robust.[257] I thus used a Matlab built-in function, *ranksum*, to perform the μ-test, and the hypothesis results and prominent p-values are summarized in **Table 3.1**. As shown in this table, the null hypothesis $H = 0$ is valid for all the 2×4 sets of measurement data (from devices A and B at four different locations), showing the strong inter-device repeatability of my c-Air platform.

		Class-100 Clean Room	*Class-1000 Clean Room*	*Indoor*	*Outdoor*
Total (count/L)	H	0	0	0	0
	P-value	0.07	0.46	0.13	0.53
PM10 (count/L)	H	0	0	0	0
	P-value	0.07	0.46	0.21	0.53
PM2.5 (count/L)	H	0	0	0	0
	P-value	0.09	0.43	0.16	0.25

Table 3.1 Inter-device repeatability using μ-test.

c-Air measurements showing forest fire influence at extended distances

On July 22, 2016, the so-called Sand Fire struck near the Santa Clarita region in California and remained uncontained for several days.[258,259] **Figure 3.5(a)** marks both the locations of this wild fire and the UCLA campus. Although UCLA was more than 40 km from the location of the fire, on July 23 around noon, smoke and ashes filled the sky near UCLA, as shown in the photo of **Figure 3.5(b)**. We obtained six air samples using the c-Air device at an outdoor environment at UCLA, as described in the above sub-section. We compared the results with a previous sample obtained on a typical day, July 7, 2016, using the same device and at the same location. The data of both days contained six 30-s air samples measured with c-Air, with a ~2-min interval between the successive samples. For each day, the particle size distributions of the six samples were averaged and the standard deviations were plotted as the histogram in **Figure 3.5(c)**. The results showed that the outdoor PM density significantly increased on the day of the wild fire, especially for particles smaller than 4 μm, which showed an approximately 80% increase. This increase in the density of smaller particles is natural because comparatively smaller particles can travel this long distance (> 40 km) and still have significant concentrations in air.

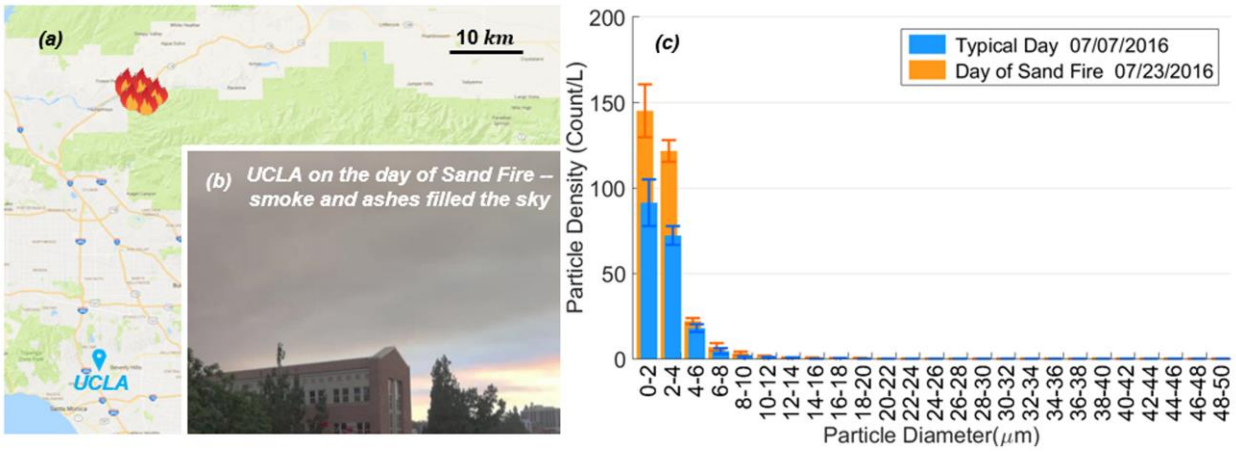

Figure 3.5 Particle size and density of the UCLA outdoor environment affected by the Sand Fire on 07/22/2016. (a) Map showing the area struck by the Sand Fire on 07/22/2016–07/23/2016. UCLA is also pinpointed on the map; it is more than 40 km from the fire location. (b) A photograph taken at ~5:00 p.m. from the second floor patio of the UCLA Engineering IV building. (c) Histogram comparison of the sample measured on a regular day, 07/07 at ~4:00 p.m., and the day of the Sand Fire, 07/22 at ~5:00 p.m., at the same location using c-Air. Each histogram bar plot with a standard deviation bar was generated from six 30-s samples. The background map of (a) is cropped from Ref. [260]. Reprinted from Ref. [23]

On September 1, 2017, another forest fire struck La Tuna Canyon Park in the greater Los Angeles area, California, which was later known as the La Tuna Fire.[261,262] As this forest fire progressed, I used our c-Air device and took a series of measurements at different geographical locations. The image processing result of the c-Air measurements shows that a significant portion of the aerosols that the La Tuna Fire generated have comparatively small size, high absorption, and noncircular shapes, which suggests that they are irregularly shaped soot-type black carbon from a hot-burning and flaming event.[263] **Figure 3.6** plots the particle sizing results. Particle sizing was performed at three spatial locations, namely, location 1 (9042 Wildwood Avenue), location 2 (9760 La Tuna Canyon Road), and location 3 (9078 La Tuna Canyon Road), as shown in **Figure 3.6 (a)**, using two platforms: a commercial optical particle counter (OPC) and c-Air. The OPC samples the air through a small channel at 2.83 L/min. From **Figure 3.6 (f–**

h), we see that with decreasing distance to the fire site (especially at location 3), the total sampled PM concentration increases, and most of the particles contributing to this increase are particles smaller than 2 µm, the count density of which is more than doubled. A similar trend is also observed by the commercial OPC, which shows an increase in particles with sizes of 1.0–2.5 µm and a significant increase in particles with sizes of 0.5–1.0 µm [as shown in **Figure 3.6 (c-e)**]. This is consistent with previously reported laboratory test results, where open-flame-generated black-carbon agglomerates (soot) were usually smaller than geologically produced dust,[264,265] with a typical size range of 20–2000 nm.[266,267]

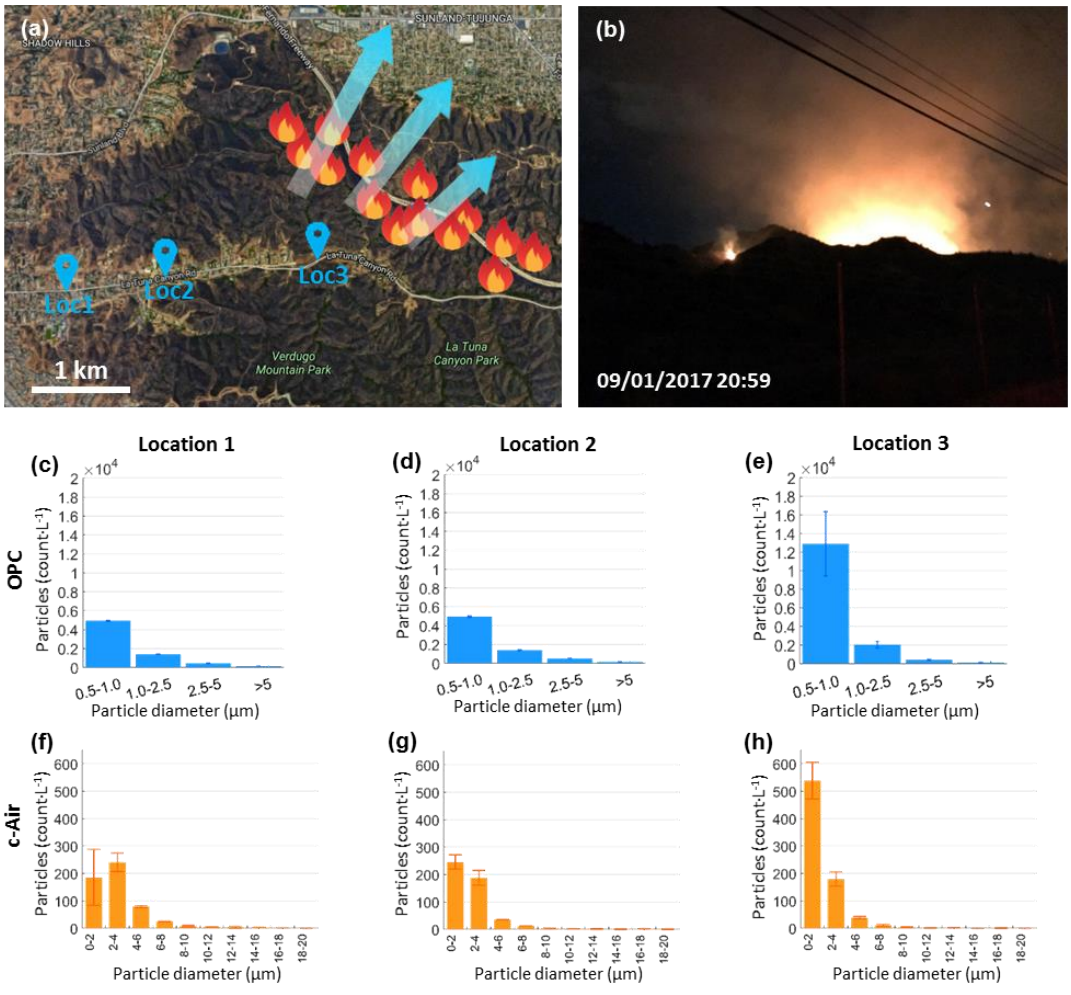

Figure 3.6 Particle sizing results at different locations. (a) Map of measurement locations during the La Tuna Fire on 09/01/2017. The blue arrows mark the wind direction. (b) Photograph of forest fire on top of a mountain that was

witnessed remotely at measurement location 3. (c–e) Particle sizing results at locations 1–3 using a commercial OPC. (f–h) Particle sizing results at the same locations using c-Air device. A significant increase is observed, especially in smaller particles, with decreasing distance to the fire site. Reprinted from Ref. [263]

Comparison of c-Air with a standard BAM PM2.5 instrument

On August 16, 2016, I employed a c-Air device at the Reseda Air Quality Monitoring Station (18330 Gault St., Reseda, CA, USA) and obtained a series of measurements during a 15-h period starting from 6:00 a.m. I compared the performance of the c-Air device with that of the conventional EPA-approved BAM PM2.5 measurement instrument (BAM-1020, Met One Instruments, Inc.). The EPA-approved BAM-1020 pumps air at $\sim 16.7\ \mathrm{L/min}$ and has a rotating filter amid airflow that accumulates PM2.5 to be measured each hour. A beta-particle source and detector pair inside measures the attenuation induced by the accumulated PM2.5 on the filter and converts it to total mass using the Beer–Lambert law. The quantity reported from BAM-1020 is hourly averaged PM2.5 mass density in $\mu g/m^3$. In comparison, the c-Air device is programmed to obtain a 30-s average particle count per 6.5 L of air volume. It performs sizing and concentration measurements using optical microscopic imaging (see Methods for details).

To enable a fair comparison, I obtained four 30-s measurements each hour, with 10- to 15-min intervals between consecutive c-Air measurements. I measured the PM2.5 densities corresponding to these samples and obtained their average as the final measured PM2.5 density for a given hour. This c-Air average was compared to the hourly average PM2.5 mass density measured by BAM-1020. The measurements of the c-Air device were obtained on the roof of the Reseda Air Sampling Station close to the inlet nozzle of BAM-1020; however, it was situated ~2 m from it to avoid interference between the two systems.

Figure 3.7 plots the comparison of the measurement results from c-Air and BAM-1020.

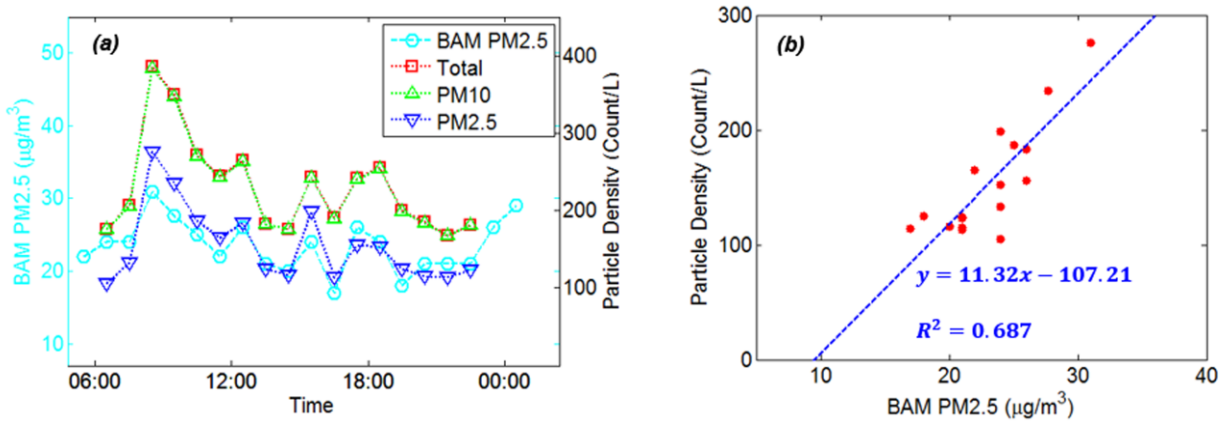

Figure 3.7 Comparison of c-Air results against a standard BAM PM2.5 instrument. (a) Superimposed hourly plot of (right axis) particle density counted by a c-Air device, and (left axes) hourly accumulated PM2.5 total mass density measured by the standard BAM PM2.5 instrument at the Reseda Air Sampling Station. (b) Linear correlation plot of PM2.5 hourly average count density measured by c-Air (y-axis) with a PM2.5 hourly average mass density measured by BAM-1020 PM2.5 measurement device (x-axis). The BAM-PM2.5 data were downloaded from Ref. [268]. Reprinted from Ref. [23]

As shown in **Figure 3.7 (c)**, c-Air's hourly average PM2.5 count density result (blue curve) closely follows the same trend as the EPA-approved BAM PM2.5 mass density result (cyan curve). I also plotted hourly averaged TSP (red curve) and PM10 (green curve) in **Figure 3.7 (a)**, which follows a similar trend as PM2.5. Lastly, I found a linear correlation between the BAM PM2.5 measurements and c-Air PM2.5 count density measurements, as shown in **Figure 3.7 (b)**, where the x-axis is the PM2.5 mass density in $\mu g/m^3$ measured by the BAM-1020, and the y-axis is the PM2.5 count density in counts/L measured by the c-Air device. Some of the variations between the two techniques may be due to several reasons: (1) Each PM2.5 particle may have a different weight; therefore, the PM2.5 count density may not directly convert to mass density of the particles; (2) There may be some air-quality variations within each hour; thus, our four 30-s measurements may not accurately represent the whole hourly average reported by

BAM-1020; (3) The cut-off size of our impactor is ~1.4 μm, which means particles smaller than this size may not be efficiently counted by our device, whereas they are counted by BAM-1020. Note also that, in **Figure 3.7 (a)** at 7:00 to 9:00 a.m., the original measurements by BAM-1020 are missing on account of the replacement of the rotating filter, which is required for the instrument's maintenance. Instead, these data points are replaced by the average of the 7:00 to 9:00 a.m. time windows on Fridays, which were measured within the same month (see Ref. [268]).

Spatial-temporal mapping of air-quality near LAX

On September 6, 2016, I employed two c-Air devices, device A and device B, to measure the spatio-temporal air quality changes around LAX. To this end, I obtained two 24-h measurements spanning two different routes that represent the longitudinal and latitudinal directions, centered at LAX (see **Figure 3.8** and **Figure 3.9**). I selected six locations in each route and performed measurements with 2h interval in each route over 24h. These raw c-Air measurements were sent to server for automatic processing to generate the particle size statistics at each time and location.

Route 1 extended from LAX to the east in a longitudinal direction, as shown in **Figure 3.8 (a)**. Along this route, I selected six sites that were located at 3.37 km, 4.34 km, 5.91 km, 7.61 km, 9.95 km, and 13.1 km east of LAX, respectively. LAX shows a pattern of a large number of flights throughout the day (7 a.m. to 11 p.m.); however, it shows a prominent valley at late night (~2 a.m.), where the number of flights is minimal, as shown by the cyan curves in **Figure 3.8 (a)**. As depicted in bubble boxes (1) to (6) in **Figure 3.8 (a)**, the c-Air measurement results of both PM2.5 and PM10 also show such a valley during late night, which illustrates the relationship between the total number of flights at LAX and the nearby PM pollutants. As the distance increases from LAX, this modulation weakens.

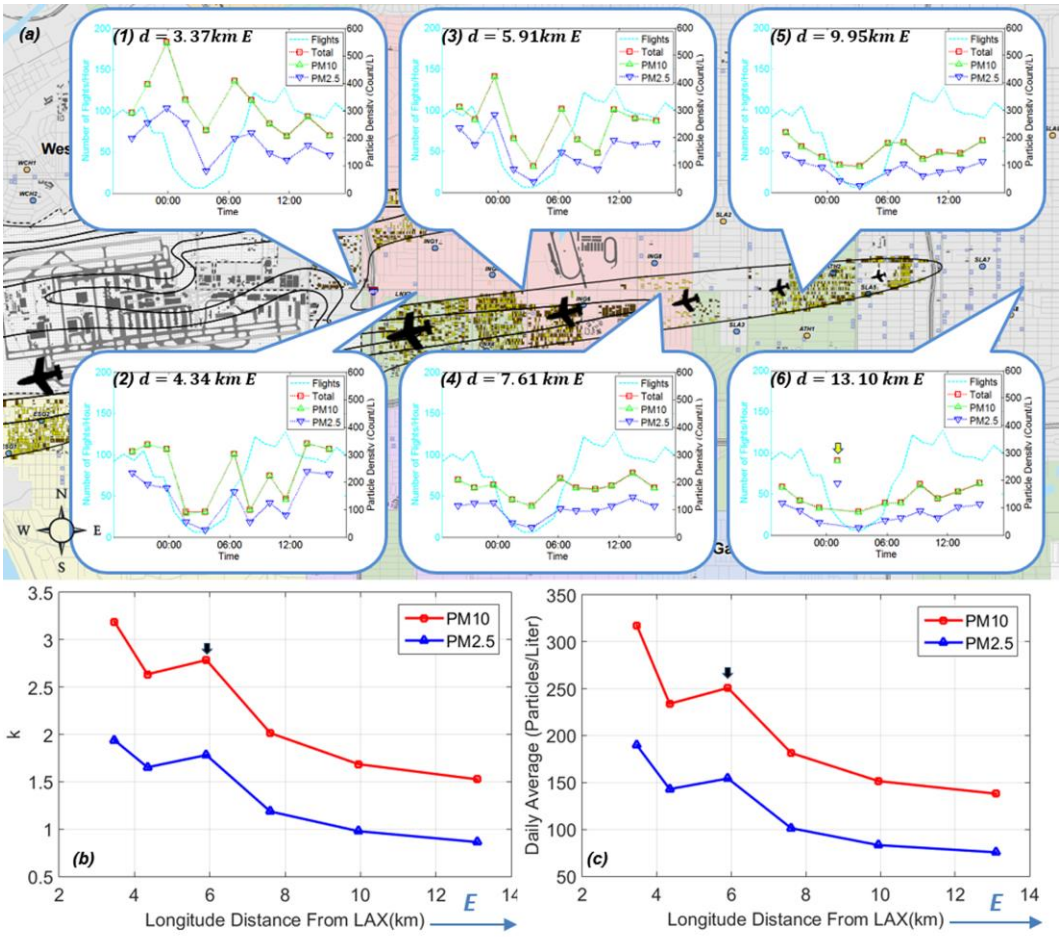

Figure 3.8 LAX measurements in the longitudinal direction using c-Air. (a) Noise map near LAX. The east side of LAX is where airplanes arrive; it is marked by tiny airplane icons. We obtained a 24-h PM measurement on 09/06/2016–09/07/2016 at the following locations: (1) 5223 W. Century Blvd., (2) 10098 S. Inglewood Ave., (3) 4011 W. Century Blvd., (4) 3000 W. Century Blvd., (5) 1407 W. 101st St. and (6) 9919 S. Avalon Blvd., as marked on (a). The fourth time point at location (a-6) was excluded from the curve because there was a large water sprinkler turned on during the measurement, which affected the c-Air performance. (b) Correlation slope plotted for locations (1–6) as a function of their longitudinal distances from LAX. (c) Daily average PM plotted for locations (1–6) as a function of their longitudinal distances from LAX. The third point in (b)–(c), as marked by a black arrow, is inconsistent with the trend, which we believe is on account of the presence of an immense (~3,400 spaces) parking lot nearby that specific measurement location. (a) depicts the noise maps of the second quarter of 2016 near LAX cropped from Ref. [269]. The total number of flights, represented by the cyan curve of (a-1) to (a-6), is plotted from the data given by Ref. [270]. Reprinted from Ref. [23].

To quantify this correlation, I defined two measures: (*1*) the correlation slope, which is the slope of the linear mapping from the total number of flights to the PM10 or PM2.5 count density (plotted as a function of the longitudinal distance from LAX in **Figure 3.8 (b)**), and (*2*) the daily average PM measured by c-Air, which is the 24-h average PM10 or PM2.5 count density for each location (also plotted as a function of the longitudinal distance from LAX in **Figure 3.8 (c)**). These figures show an exponential trend for both the correlation slope and the daily average PM as a function of the distance from LAX. Moreover, they indicate that the impact of the airport in increasing air pollution is significant, even >7 km from its location. This has also been independently confirmed in earlier studies,[271] using a commercially available optical scattering based PM detection technology[251] that has a limited dynamic range of particle size and density, and more than an order of magnitude lower throughput compared to ours due to its serial read-out scheme as discussed in our Introduction section.

Unlike Route 1, Route 2 extended from the south to the north of LAX, spanning a latitudinal direction. The six locations chosen in this route were 3.58 km, 1.91 km, 0.50 km, 0.01 km, −1.46 km, and −2.19 km north of LAX, respectively. Similar to Route 1, bubble boxes (1) to (6) of **Figure 3.9** plot the time variations of the PM concentration during the sampling points corresponding to Route 2. These results once again show a similar trend of PM variation in accordance with the total number of flights at LAX. Similar to Route 1, as the measurement location distance from LAX increases, the modulation strength diminishes. The above finding is supported by the correlation slope shown in **Figure 3.9(b)** and the daily average PM shown in **Figure 3.9(c)**, both of which are a function of the latitudinal distance north of LAX. We note that the decrease is more rapid in this latitudinal direction (south-to-north, Route 2) than the

longitudinal direction (west-to-east, Route 1), which may be on account of the typical west winds near LAX during the summer, which cause the particles to travel longer distances in air.[272]

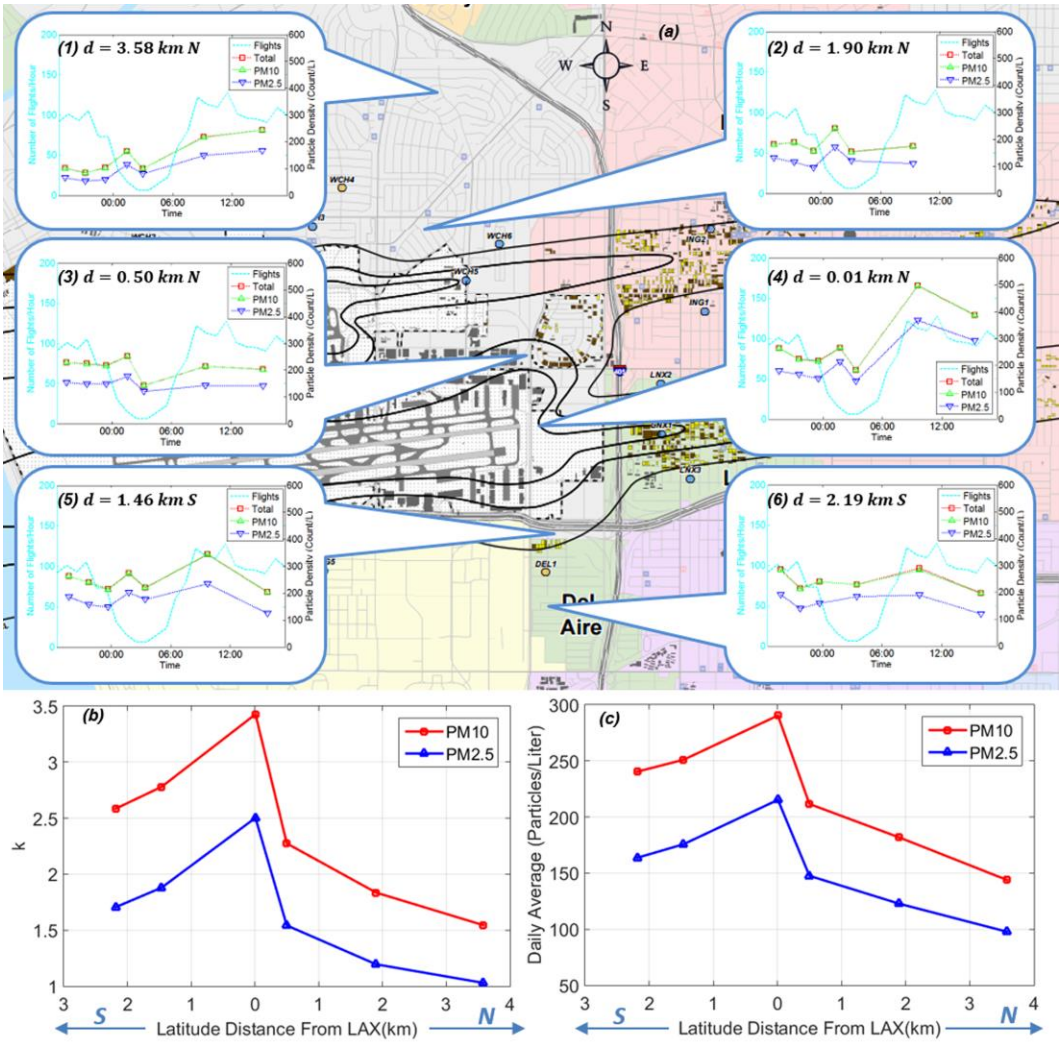

Figure 3.9 LAX measurements in the latitudinal direction using c-Air. (a) Noise map near LAX. We obtained a 24-h PM measurement on 09/06/2016–09/07/2016 at locations (1) 6076 W. 76th St., (2) 8701 Airlane Ave., (3) 5625 W. Century Blvd., (4) 10400 Aviation Blvd., (5) 5457 W. 117th St., and (6) 5502 W. 122nd St., as marked on (a). (b) Correlation slope plotted for locations (1–6) as a function of their latitudinal distances from LAX. (c) Daily average PM plotted for locations (1–6) as a function of their latitudinal distances from LAX. (a) is the noise map of the second quarter of 2016 near LAX cropped from Ref. [269]. Reprinted from Ref. [23].

3.1.3 Conclusion

In this section, I demonstrated a portable and cost-effective PM imaging, sizing, and quantification platform, called c-Air, which uses lens-less computational microscopy and machine learning. The platform consists of a field-portable device weighing approximately 590 g, a smartphone app for device control and display of results, and a remote server for automated processing of digital holographic microscope images for PM measurements based on a custom-developed machine learning algorithm. The performance of the device was validated by measuring air quality at various indoor and outdoor locations, including an EPA-regulated air sampling station, where I compared c-Air results with those of an EPA-approved BAM device, and a close correlation was shown. I further used c-Air platform for spatio-temporal mapping of air-quality near LAX, which showed the PM concentration varying throughout the day in accordance with the total number of flights at LAX. The strength of this correlation, as well as the daily average PM, exponentially decreased as a function of the increasing distance from LAX. The c-Air platform, with its microscopic imaging and machine learning interface, can have a wide range of applications in air quality regulation and improvement.

3.1.4 Methods

Impaction-based air-sampler

To capture aerosols, I employ an impaction-based air sampler (i.e., an impactor) on account of its high-throughput, simple hardware, and compatibility with microscopic inspection.[273] As shown in **Figure 3.1(d)**, the impactor consists of an impaction nozzle and a sticky sampling coverslip (Air-O-Cell Sampling Cassette, Zefon International, Inc.). A pump drives the laminar airstream through the nozzle at high speed. The sticky coverslip is placed to directly face the airstream. The airstream can be easily redirected while the aerosol inside the stream impacts with

and is collected by the sticky coverslip. This collection is subsequently used for computational imaging. The aerosol capture by the impactor is actually a random process. The probability that an individual aerosol particle passing through the impactor will be captured depends on the particle size, laminar airflow rate, and nozzle width. This probability is related to the Stokes number (*Stk*):[252]

$$Stk = \frac{\rho_P d_P^2 U C_c}{9\eta D_j} \tag{3.2}$$

where ρ_P is the particle mass density, d_P denotes the particle diameter, U represents the flow velocity, η is the air viscosity, D_j denotes the nozzle diameter, and C_c is the slip correction coefficient. The impaction efficiency increases as Stk increases. Based on the same terminology, the cut-off size, d_{50}, is defined as the diameter of the particle at which the impaction efficiency decreases to 50%. In my experimental design, the air sampler (with a nozzle of 1.1 mm by 14.5 mm) was connected to and driven by a miniaturized pump (M00198, GTEK Automation) with a throughput of 13 L/min. Based on the above relationship, the 50% cut-off sampling diameter of the c-Air can be estimated as $d_{50} = 1.4$ μm.[274]

c-Air lens-less on-chip microscope and air sampling design

For rapid imaging and inspection of the collected aerosols, I combined the impactor with a lens-less microscope, as shown in **Figure 3.1**. Similar to a typical lens-less imaging setup, the sticky coverslip, which was the sample to be imaged, together with the impactor nozzle cartridge, was directly placed on a color complementary metal-oxide semiconductor (CMOS) image sensor (OV5647, 5 MPix, 1.4 μm pixel size) at a distance of approximately 400 μm from the image sensor to the sticky coverslip. Three fiber-coupled light-emitting diodes (LEDs) (red: 624 nm; green: 527 nm; blue: 470 nm) were fixed on top of the device sampler tube at

approximately 8 cm from the image sensor chip. Using LED illumination, the aerosol samples, which were captured by the sticky coverslip, casted in-line holograms. These holograms were recorded by the CMOS image sensor for holographic reconstruction and further processing to determine the PM statistics.

The image capturing and air sampling processes are illustrated in **Figure 3.2(b)**. When the user hits "capture a sample" from the app interface, the three LEDs are sequentially turned on/off, and three background images are thereby obtained. These background images depict the state of the air-sampler surface prior to the intake of the air to be sampled. Next, the pump was powered on to push the air through the sampler for 30 s, thereby screening 6.5 L of air. The three LEDs were then sequentially turned on/off, and three sample images were thereby obtained with the newly captured aerosol particles. These background images and sample images were both synced to the server for further processing. After subtraction of the sample image from its corresponding background image, a differential hologram was formed, which contained the information of only the newly captured particles. For particle sizing, I used only the images captured under the green LED illumination. By merging all the differential holographic images captured using the three LEDs, red–green–blue color images of the captured particles could also be obtained, revealing the color information of the specimen, if desired. To avoid awaiting completion of the steps before a new sample could be obtained, I programmed the sampling process in a parallel way. Accordingly, when a new sampling request arrived before the previous result was synced, the un-synced sample was cached first. It was later synced when the device became idle. The entire device sampling process was controlled by a custom-developed program on a microcomputer (Raspberry Pi A+), along with a custom-designed circuit.

Particle detection using digital "peeling"

Direct back-propagation of the acquired hologram (using Equation (1.17)) to the auto-focused sample plane generates a spatial artifact, called the twin-image noise. This twin-image artifact affects the detection of aerosols. To address this problem, I employ an iterative particle peeling algorithm[8] in holographic reconstruction process, combined with a support vector machine (SVM)-based learning model to further reject these spatial artifacts. This peeling algorithm contains four cycles of detection and erasing ("peeling out") of the particles at progressively increasing thresholds, i.e., 0.75, 0.85, 0.92, and 0.97, where the background is centered at 1.0 during the differential imaging process. The highest threshold (0.97) is selected as 3σ from the background mean, where $\sigma \approx 0.01$ is the standard deviation of the background. I use a morphological reconstruction process[104] to generate the image mask instead of using a simple threshold. Because most particles have a darker center and a somewhat weaker boundary, using a single threshold may mask parts of the particle, potentially causing the particle to be missed or re-detected multiple times in subsequent peeling cycles. This is avoided by using a morphological reconstruction process.

In each cycle of this digital particle peeling process, I first adjust the image focus using the auto-focusing algorithm (see section 1.5). Then, the morphological reconstruction is employed to generate a binary mask, where each masked area contains a particle. For each mask, I crop a small image (100 × 100 pixels) and perform fine auto-focusing on this small image to find the correct focus plane of the corresponding particle. At this focus plane, I extract various spatial features of the particle, e.g., minimum intensity I_m, average intensity I_a, area A, and maximum phase θ_M. We then propagate the image to five different planes uniformly spaced between 20 μm above and 20 μm below this focus plane. The Tamura coefficient[62] of this focus plane is

calculated and compared to the coefficients of these five other planes. The ratio of the Tamura coefficient at this focus plane against the highest Tamura coefficient of all six planes is defined as another feature, R_{Tam}. These four features, I_m, θ_M, A, and R_{Tam}, are then fed into an SVM-based learning model to digitally separate spatial artifacts from true particles and reject such artifacts. This learning algorithm is detailed in the following sub-section. After all the detected particles in this peeling cycle are processed, I digitally peel out these "counted" particles, i.e., replace the thresholded area corresponding to each detected particle with the background mean, on both the image and twin image planes. I then move to the next peeling cycle with a higher threshold and repeat the same steps.

After completing all four peeling cycles, the extracted features, I_m, θ_M, and A, are further utilized for particle sizing using a machine-learning algorithm, as detailed further below. This sizing process is only performed on true particles, which generates a histogram of particle sizes and density distributions, as well as various other parameters, including, for example, TSP, PM10, and PM2.5, reported as part of c-Air result summary.

Elimination of false-positives using a trained support vector machine

To further avoid false-positives in detecting aerosols, I used a trained linear SVM that is based on four features, I_m, θ_M, A, and R_{Tam}, as described in the previous sub-section, to distinguish spatial artifacts from true particles and increase c-Air detection accuracy. These spatial features were selected to provide the best separation between the true- and false-particles. To train this model, I obtained two air sample images using a c-Air prototype, one indoor and one outdoor. Then, in addition to our c-Air based analysis, I physically extracted the sampling coverslip and inspected the captured particles under a benchtop bright-field microscope using a 40X objective lens. I compared the thresholded areas in peeling cycle and lens-less reconstruction process with

the images of the benchtop microscope to mark each one of these detected areas as a true particle or a false one. Using this comparison, I labeled a total of more than 2,000 thresholded areas and fed half of this training data into our SVM model[275] (implemented in Matlab using the function "svmtrain"). The other half was used for blind testing of the model, which showed a precision of 0.95 and a recall of 0.98.

Machine learning based particle detection and sizing

I used a custom-designed machine-learning algorithm trained on size-calibrated particles to obtain a mapping from the detected spatial characteristics of the particles to their diameter. For this purpose, I used some spatial features extracted from the holographic particle images, including e.g., minimum intensity I_m, average intensity I_a, and area A, and developed a second-order regression model that maps these features to the sizes of the detected particles in microns. The model is deterministically learned from size-labeled particle images, which are manually sized using a standard benchtop microscope. Specifically, after I extract features I_m, I_a, and A of the masked region in a particle peeling cycle, I strive to find a model, f, that maps these features to the particle diameter D (in microns), i.e.,

$$\mathrm{D} = \mathrm{f}\left(\mathrm{I_m}, \mathrm{I_a}, \sqrt{\mathrm{A}}\right) \tag{3.3}$$

where f can potentially have infinite dimensions. However, I employ a simplified second-order polynomial model of f and extend the features to the second-order by defining. Then, I define a linear mapping, θ, that relates the second-order features to the diameter of the particle:

$$\mathrm{X} = [1, \mathrm{I_m}, \mathrm{I_a}, \sqrt{\mathrm{A}}, \mathrm{I_m^2}, \mathrm{I_a^2}, \mathrm{A}, \mathrm{I_m I_a}, \mathrm{I_m}\sqrt{\mathrm{A}}, \ \mathrm{I_a}\sqrt{\mathrm{A}}] \tag{3.4}$$

$$\mathrm{D} = \mathrm{f}\left(\mathrm{I_m}, \mathrm{I_a}, \sqrt{\mathrm{A}}\right) = \theta^{\mathrm{T}}\widehat{\mathrm{X}} = \theta^{\mathrm{T}}\left(\frac{\mathrm{X} - \mu}{\sigma}\right) \tag{3.5}$$

where ^{T} refers to the transpose operation, and μ and σ represent the mean and standard deviation of X, respectively. Based on the above mapping, I used 395 size-labeled microbeads for training and blind testing. These polystyrene microbeads ranged in diameter from ~1 μm to ~40 μm, as shown in **Figure 3.3**. The ground truth sizes of these particles were manually measured under a benchtop microscope with a 100×/0.95NA objective lens. The same samples were additionally imaged using the c-Air platform to obtain the corresponding lens-less images and extract spatial features, I_m, θ_M, and A. For training the machine learning model, I first randomly and evenly separated the microbead samples into respective training and testing sets. After extending the features to the second-order (equation (3.4)) and performing normalization (equation (3.5)), I fitted the parameter vector θ by minimizing the difference between the training feature mapping ($\theta^T \hat{X}_{tr}$) and the calibrated diameter (D_{tr}^{mic}) i.e.,

$$\min_{\theta} \left\| \theta^{\mathrm{T}} \hat{\mathrm{X}}_{\mathrm{tr}} - \mathrm{D}_{\mathrm{tr}}^{\mathrm{mic}} \right\|_1 \tag{3.6}$$

This minimization was performed by CVX, a software package designed for solving convex optimization problems.[276] The same trained parameter was then applied for the cross-validation set, which was comprised of another 197 microbeads. Particle sizing training errors (E_{tr}) and testing errors (E_{cv}) were validated by evaluating the norm of difference:

$$\mathrm{E}_{\mathrm{tr}} = \left\| \theta^{\mathrm{T}} \hat{\mathrm{X}}_{\mathrm{tr}} - \mathrm{D}_{\mathrm{tr}}^{\mathrm{mic}} \right\|_{\mathrm{p}} \tag{3.7}$$

$$\mathrm{E}_{\mathrm{cv}} = \left\| \theta^{\mathrm{T}} \hat{\mathrm{X}}_{\mathrm{cv}} - \mathrm{D}_{\mathrm{cv}}^{\mathrm{mic}} \right\|_{\mathrm{p}} \tag{3.8}$$

where $\theta^T \hat{X}_{cv}$ is the testing feature mapping, and D_{cv}^{mic} is the calibrated diameter for the testing set. In addition, $p = 1$ is used for calculating the "mean error," and $\mathrm{p} = \infty$ is used for calculating the "maximum error." Note that this training process only needs to be done once, and the trained parameter vector, θ, and the normalization parameters, μ and σ, are then saved and

subsequently used for *blind* particle sizing of all the captured aerosol samples using c-Air devices.

Converting particle count to particle density

For each sample (and the corresponding c-Air differential hologram), the particle detection and sizing algorithm, as described in the main text, provides the particle count in the number of particles for different sizes/diameters. To facilitate a more universal unit, I convert the sampling particle count, N_i, to particle density n_i (in $\mathrm{counts/L}$) using the following equation:

$$\mathrm{n_i} = \frac{\mathrm{N_i}}{\mathrm{Q \cdot t}} \cdot \frac{\mathrm{L_{sensor}}}{\mathrm{L_{total}}} \tag{3.9}$$

where $Q = 13\ \mathrm{L/min}$ is the flow rate of air, and $t = 0.5\ \mathrm{min}$ is the typical sampling duration. In addition, $L_{total} = 15.5\ \mathrm{mm}$ is the total length of the impactor nozzle slit, and $L_{sensor} = 3.67\ \mathrm{mm}$ is the part of the slit being imaged, which equals the longer edge of the CMOS sensor active area. The conversion equation here assumes that the particle distribution is uniform along the sampler nozzle length, which is a valid assumption because the nozzle tapering is in its orthogonal direction, while the structure of the impactor in length direction is spatially invariant.

3.2 Label-Free Bioaerosol Sensing

Conventional bio-aerosol sensing requires the sampled aerosols in the field to be transferred to a laboratory for manual inspection, which can be rather costly and slow, also requiring a professional for labeling and microscopic examination of the samples. Here, built upon my previous work of c-Air described in section 3.1, I demonstrate label-free bio-aerosol sensing using a field-portable and cost-effective device based on holographic microscopy and deep-learning, which screens bio-aerosols at a throughput of 13 L/min. Two different deep neural

networks are designed to rapidly reconstruct the amplitude and phase images of the captured bio-aerosols, and to classify the type of each bio-aerosol that is imaged. As a proof-of-concept, I studied label-free sensing of common bio-aerosol types, e.g., Bermuda grass pollen, oak tree pollen, ragweed pollen, Aspergillus spore, and Alternaria spore and achieved >94% classification accuracy. The presented label-free bio-aerosol measurement device, with its mobility and cost-effectiveness, will find several applications in indoor and outdoor air quality monitoring.

3.2.1 Introduction

A human adult inhales about seven liters of air every minute, which on average contains 10^2-10^3 micro-biological cells (bio-aerosols) [277,278]. In some contaminated environments, this number can easily exceed 10^6 [277,279–281]. These bio-aerosols are micro-scale airborne living organisms that originate from plants or animals, and include pollens, mold/fungi spores, bacteria, and viruses. They are generated both naturally and anthropogenically, from e.g. animal houses [282,283], composting facilities [282,283], and construction sites [283]. These bio-aerosols can stay suspended in the air for prolonged periods of time [282], remain at significant concentrations even far away from the generating site (up to one kilometer) [284–286], and can travel through continental distances [287]. Basic environmental conditions, such as temperature and moisture level, can also considerably influence bio-aerosol formation and dispersion [283,288]. Inhaled by a human, they can stay in the respiratory tract and cause irritation, allergies, various diseases including cancer and even pre-mature death [247,277,278,282,283,286,287,289,290]. In fact, bio-aerosols account for 5-34% of indoor particulate matter (PM) [291]. In recent years, there has been increased interest in monitoring environmental bio-aerosols, and understanding their composition, to avoid and/or mitigate their negative impacts on human health [277,283,292,293], in both peacetime and in threat of biological attacks [290].

Currently, most of the bio-aerosol monitoring activities still rely on a technology that was developed more than fifty years ago [283,294,295]. In this method, an aerosol sample is taken in the inspection site using a sampling device such as an impactor, a cyclone, a filter, or a spore trap. This sample is then transmitted to a laboratory, where the aerosols are transferred to certain liquid media or solid substrates and inspected manually under a microscope or through culture experiments. The microscopic inspection of the sample usually involves labeling through a colorimetric or fluorescence stain to increase the contrast of the captured bio-aerosols under a microscope [296,297]. Regardless of the specific method that is employed, the use of manual inspection in a laboratory, following a field collection, significantly increases the costs and delays the reporting time of the results. Partially due to these limitations, out of ~10,000 air-sampling stations worldwide, only a very small portion of them have bio-aerosol sensing/measurement capability. Even in developed countries, bio-aerosol levels are only reported on a daily basis at city scales [298]. As a result, human exposure to bio-aerosols is hard to quantify with the existing set of technologies.

Driven by this need, different techniques have been emerging towards potentially label-free, on-site and/or real-time bio-aerosol monitoring. In one of these techniques, the air is driven through a small channel, and an ultraviolet (UV) source is focused on a nozzle of this channel, exciting the auto-fluorescence of each individual bio-aerosol flowing through the nozzle [299–303]. This auto-fluorescence signal is then captured by one or more photodetectors, used to differentiate bio-aerosols from non-fluorescent background aerosols. Recently, other machine learning algorithms have also been applied to classify bio-aerosols from their auto-fluorescence signals[302,303]. However, measuring auto-fluorescence may not provide sufficient specificity towards classification. To detect weak auto-fluorescence signals, this design also requires strong

UV sources, sensitive photodetectors and high-performance optical components, making the system relatively costly and bulky. Furthermore, the sequential read-out scheme in these flow-based designs also limits their sampling rate and throughput to < 5 L/min. Alternative bio-aerosol detection methods rely on anti-bodies to specifically capture bio-aerosols of interest on e.g., a vibrational cantilever [304,305], or a surface plasmon resonance (SPR) substrate [292,306], which can then detect these captured bio-aerosols through a change in the cantilever vibrational frequency or a shift in the SPR spectrum, respectively. These approaches provide very sensitive detection of a specific type of bio-aerosols. However, their performance can be compromised by non-specific binding and/or changes in the environmental conditions (e.g., temperature, moisture level etc.), impacting the effectiveness of the surface chemistry. Moreover, the reliance to specific antibodies makes it harder for these approaches to scale up the number of target bio-aerosols and cover unknown targets. Bio-aerosol detection and composition analysis using Raman spectroscopy has also been demonstrated [307,308]. However, due to weaker signal levels and contamination from background spectra, the sensitivities of these methods have been relatively low despite their expensive and bulky hardware; it is challenging to analyze or detect e.g., a single bio-aerosol within a mixture of other aerosols. It is also possible to detect bio-aerosols by detecting their genetic material (e.g., DNA), using polymerase chain reaction (PCR) [309], enzyme-linked immunosorbent assays (ELISA) [310] or metagenomics [311,312], all of which can provide high sensitivity and specificity. However, these detection methods are usually based on post-processing of bio-aerosols in laboratory environments (i.e., involves field sampling, followed by the transportation of the sample to a central laboratory for advanced processing), and are therefore low-throughput, also requiring an expert. Therefore, there is still an urgent

unmet need for accurate, label-free and automated bio-aerosol sensing to cover a wide range of bio-aerosols, ideally within a field-portable, compact and cost-effective platform.

To address this need, I developed a high-throughput and mobile bio-aerosol detection system based on c-Air device (**Figure 3.1**) and deep learning. The device design uses a combination of an impactor and a lens-less digital holographic on-chip microscope [16,23,34,187,313]: bio-aerosols in air are captured on the impactor substrate at a sampling rate of 13 L / min. These collected bio-aerosols generate diffraction holograms recorded directly by an image sensor chip that is positioned right below the substrate. Each hologram contains information of the complex optical field (amplitude and phase information) of each individual bio-aerosol. After digital holograms of bio-aerosols are acquired and transmitted to a remote server, these holograms are rapidly processed through an image-processing pipeline (**Figure 3.10**), within a minute, reconstructing the entire field-of-view (FOV) of our device, i.e., 4.04 mm^2, over which the captured bio-aerosols are analyzed. Enabled by deep convolutional neural networks (CNNs) [129], this reconstruction algorithm first reconstructs both the amplitude and phase image of each individual bio-aerosol with sub-micron resolution, and then performs automatic classification of the imaged bio-aerosols into pre-trained classes and counting the density of each class in air. As a proof-of-concept, I demonstrate the reconstruction and label-free sensing of five different types of bio-aerosols: Bermuda grass pollen, oak tree pollen, ragweed pollen, Aspergillus spore, and Alternaria spore – as well as non-biological aerosols as part of the default background pollution. The Bermuda grass, oak tree and ragweed pollens have long been recognized as some of the most common grass, tree and weed-based allergens that can cause severe allergic reactions [287,314–316]. Similarly, the Aspergillus and Alternaria spores are two of the most common mold spores found in air [277,279–281,285] and can cause allergic reactions and various diseases [282,283,286,287].

Furthermore, Aspergillus spores have been proven to be a culprit of asthma in children [282]. Some of these mold species/sub-species can also generate mycotoxins that weaken the human immune system [283]. In this work, a deep CNN is trained to differentiate these six different types of aerosols, achieving an accuracy of 94% using our mobile instrument. This label-free bio-sensing platform can be further scaled up to specifically detect other types of bio-aerosols by training it using purified populations of new target object types as long as these bio-aerosols exhibit unique spatial and/or spectral features that can be detected through our holographic imaging system.

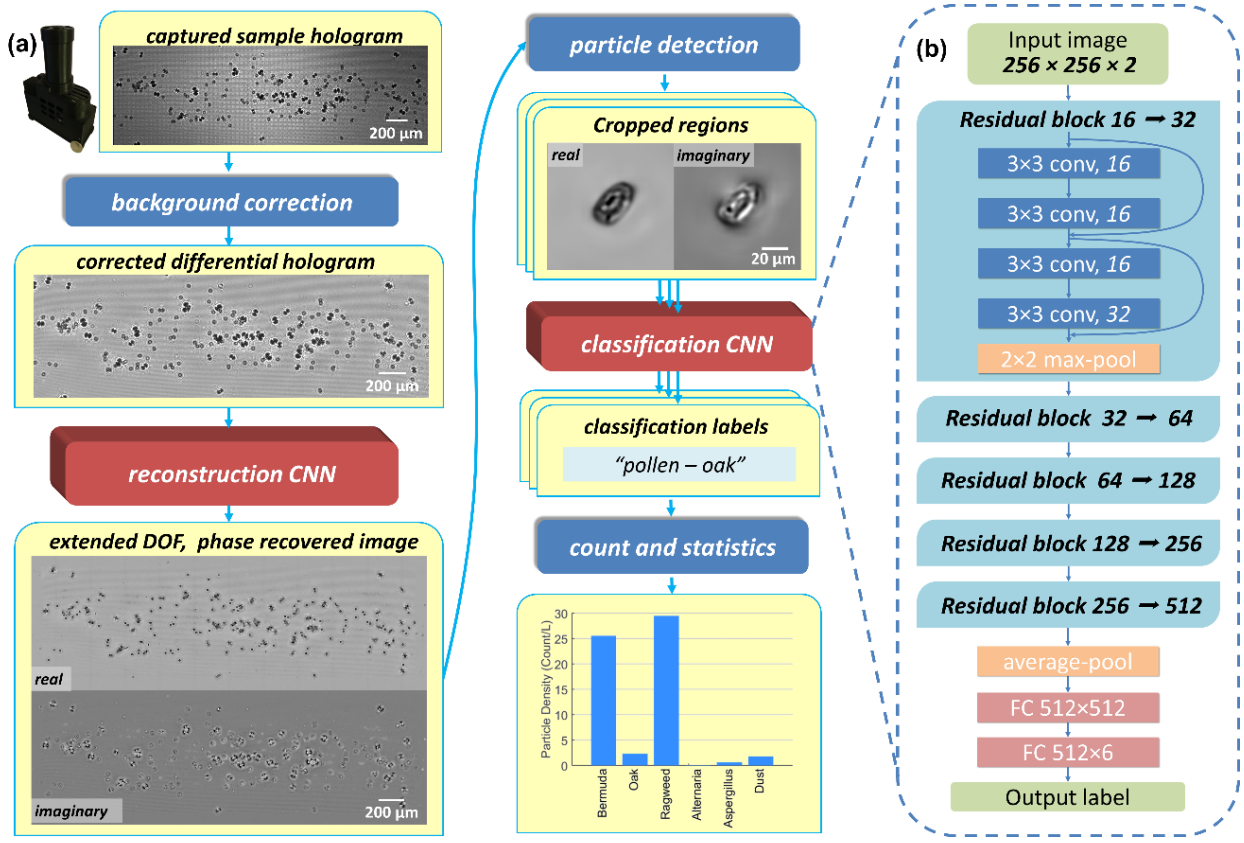

Figure 3.10 Label-free bio-aerosol sensing. (a) Image reconstruction and bio-aerosol classification work-flow. (b) Architecture of the classification CNN. conv: convolutional layer. FC: fully-connected layer. Reprinted from Ref. [24]

To the best of my knowledge, this is the first demonstration of automated label-free sensing and classification of bio-aerosols using a portable and cost-effective device, which is enabled by computational microscopy and deep-learning. Compared to earlier results on PM sensing using

mobile microscopy without any classification capability (section 3.1), this section reports label-free and automated bio-aerosol sensing using deep learning (which is used for both image reconstruction and classification), providing a unique capability for specific and sensitive detection and counting of e.g., pollen and mold particles in air. The presented platform can find a wide range of applications in label-free bio-aerosol sensing and environmental monitoring.

3.2.2 Results and discussion

Quantification of spatial resolution and field-of-view

A USAF-1951 resolution test target is used to quantify the spatial resolution of our device, where the smallest resolvable line is group nine, element one (with a line width of 0.98 μm), which in this case is limited by the pixel pitch of the image sensor chip (1.12 μm). For the current bio-aerosol sensing application, this resolution provides accurate detection performance, revealing the necessary spatial features of the particles in both the phase and amplitude image channels, as will be detailed in subsequent sub-sections. In case future applications require better spatial resolution to reveal even finer spatial structures of some target bio-aerosols, the resolution of our platform can be further improved by using an image sensor chip with a smaller pixel pitch, and/or by applying pixel super-resolution techniques that can digitally achieve an effective pixel pitch < 0.5 μm [33,48,53].

In my device design, the image sensor chip has an active area of $3.674\ mm \times 2.760\ mm = 10.14\ mm^2$, which would normally be the sample FOV for a lens-less on-chip microscope. However, the imaging FOV is smaller than this because the sampled aerosols deposit directly below the impaction nozzle, thus the active FOV of the mobile instrument is defined by the overlapping area of the image sensor chip and the impactor nozzle, which results in an effective

FOV of $3.674\,mm \times 1.1\,mm = 4.04\,mm^2$. This FOV can be further increased up to the active area of the imager chip by customizing the impactor design with a larger nozzle width.

Label-free bio-aerosol image reconstruction

For each bio-aerosol measurement, two holograms are taken (before and after sampling the air) by the device, and their per-pixel difference is calculated forming a differential hologram [26,317]. This differential hologram is numerically back-propagated in free space by an axial distance of ~750 μm to roughly reach the object plane of the sampling surface (see Methods for details). This axial propagation distance does not need to be precisely known, and in fact all the aerosols within this back-propagated image are automatically autofocused and phase recovered at the same time using a deep neural network that was trained with out-of-focus holograms of particles (within +/-100 μm of their corresponding axial position) to extend the depth-of-field (DOF) of the reconstructions [129], i.e. the HIDEF approach as described in section 2.3. This feature of the neural net is extremely beneficial to speed up auto-focusing and phase recovery steps since it reduces the computational complexity of the reconstructions from $O(n \cdot m)$ to $O(1)$, where n refers to the number of aerosols within our FOV and m refers to the axial search range that would have been used for auto-focusing each particle using classical holographic reconstruction methods. In this regard, deep learning is crucial to rapidly reconstruct and auto-focus each bio-aerosol's phase and amplitude image.

To illustrate the reconstruction performance of this method, **Figure 3.11** shows the raw holograms, back-propagation and neural network results of six different cropped region-of-interests (ROIs), one for each of the six classes used in this paper (Bermuda grass pollen, oak tree pollen, ragweed pollen, Alternaria mold spores, Aspergillus mold spores, and generic dust).

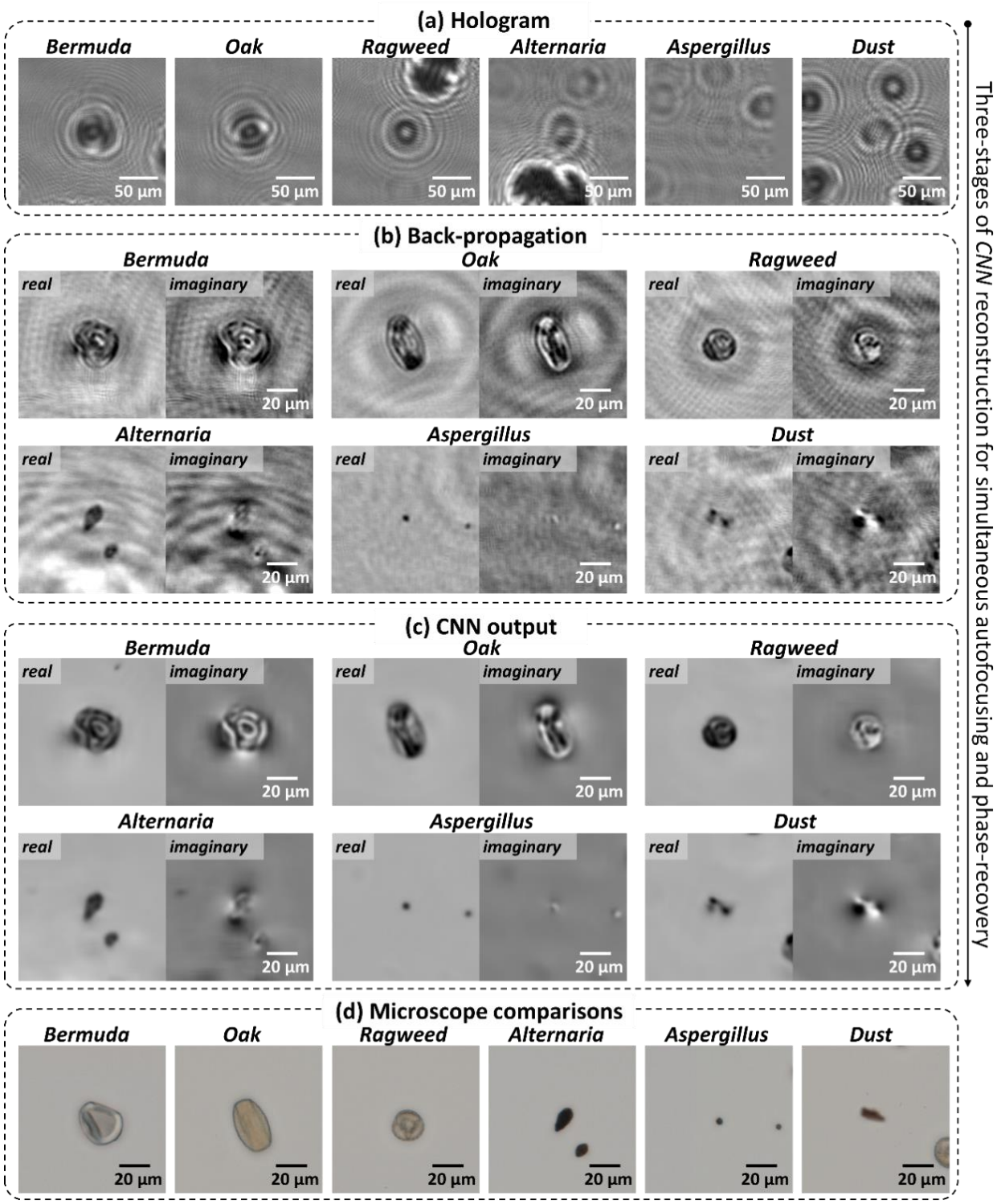

Figure 3.11 Examples of the reconstructed images of different types of bio-aerosols. (a) Cropped raw hologram. (b) Back-propagated holographic reconstruction. (c) CNN-based hologram reconstruction. (d) Corresponding regions of interest imaged by a benchtop scanning microscope with 40× magnification. Reprinted from Ref. [24]

The propagation distance (750 μm) is not exact for all these particles, which would normally result in de-focused images. This defocus is corrected automatically by my trained neural network, as shown in **Figure 3.11(c)**. In addition, the twin-image and self-interference artifacts of holographic imaging (e.g., the ripples at the background of **Figure 3.11(b)**) are also eliminated in Figure 3(c), demonstrating phase-recovery in addition to auto-focusing on each captured particle. Microscope comparisons captured under a 20×/0.75NA objective with a 2× adapter are also shown for the same six ROIs (**Figure 3.11(d)**). The neural network output (**Figure 3.11(c)**) clearly illustrates the morphological differences among these different aerosols, in both the real and imaginary channels of the reconstructed images, providing unique features for classification of these aerosols, as will be discussed in the next sub-section.

Bio-aerosol image classification

A separate CNN is developed that takes a cropped ROI and automatically assigns one of the six class labels for each detected aerosol (see **Figure 3.10** and Methods for details). **Table 3.2** reports the classification precision and recall on the testing set, as well as their harmonic mean, known as F-number (F#), which are defined as:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{3.10}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3.11}$$

$$\text{F\#} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.12}$$

As shown in **Table 3.2**, an average precision of ~94.0%, and an average recall of ~93.5% are achieved for the six labels using this classification CNN for a total number of 1,391 test particles that were imaged by my instrument. In **Table 3.2**, the classification performance of my mobile device is relatively lower for Aspergillus spores compared to other classes. This is due to the fact

that (1) Aspergillus spores are smaller in size (~ 4 μm), so their fine features may not be well-revealed under the current imaging system resolution, and (2) the Aspergillus spores sometimes cluster and may exhibit a different shape compared to an isolated spore (for which the network was trained for). In addition to these, the background dust images used in our testing are captured along the major roads with traffic. Although it should contain mostly non-biological aerosols, there is a finite chance that a few bio-aerosols may also be present in our data set, leading to mislabeling. **Table 3.2** also compares the performance of two other classification methods on the same data set, namely AlexNet [101] and support vector machine (SVM) [318]. AlexNet, although has more trainable parameters in the network design (because of the larger fully connected layers), performs ~1.8% worse in precision and 1.2% worse in recall compared to the CNN developed in this work. SVM, although very fast to compute, has significantly worse performance than our CNN models, reaching only 78.1% precision and 73.2% recall on average for our testing set. Both the SVM and AlexNet are trained and tested on the same training, validation, and testing sets, matching the CNN designed, also using a similar number of epochs (~200).

	This paper			*AlexNet*			*SVM*		
	Preci.	**Recall**	**F#**	**Preci.**	**Recall**	**F#**	**Preci.**	**Recall**	**F#**
Bermuda	0.925287	0.851852	0.887052	0.893855	0.846561	0.869565	0.769231	0.61674	0.684597
Oak	0.930464	0.975694	0.952542	0.940972	0.940972	0.940972	0.84375	0.690341	0.759375
Ragweed	0.964427	0.976	0.970179	0.931559	0.98	0.955166	0.730077	0.8875	0.801128
Alternaria	0.962963	0.962963	0.962963	0.933333	0.972222	0.952381	0.587179	0.970339	0.731629
Aspergillus	0.848485	0.937799	0.890909	0.795556	0.856459	0.824885	0.782222	0.671756	0.722793
Dust	0.944186	0.849372	0.894273	0.843602	0.74477	0.791111	0.833333	0.6	0.697674
Average	0.940059	0.934515	0.936591	0.922129	0.922511	0.921901	0.781019	0.731527	0.748367

Table 3.2 Precision and recall of our bio-aerosol classification results using a convolutional neural network. Comparison of our neural network against two other machine learning methods, AlexNet [101] and support vector machine (SVM) [318], is also provided.

Bio-aerosol mixture experiments

To further quantify the label-free sensing performance of c-Air platform, I did two additional sets of experiments – one with a mixture of the three pollens, and another with a mixture of the two mold spores. In addition, in each experiment there were also unavoidably dust particles (background PM) other than the pollens and mold spores that were introduced into our device and were sampled and imaged on the detection substrate. The sampled sticky substrate in each experiment was also examined (after lens-less imaging) under a scanning microscope with 40× magnification, where the corresponding FOV that was analyzed by c-Air was scanned and the captured bio-aerosols inside each FOV were manually labeled and counted by a microbiologist (for comparison purposes). The results of this comparison are shown in **Figure 3.12(a-f)**, where **Figure 3.12(a-d)** is from four independent pollen mixture experiments and **Figure 3.12(e,f)** is from two independent mold spores mixture experiments.

To further quantify the detection accuracy, **Figure 3.12(g-l)** plots the results of **Figure 3.12(a-f)** individually for each of the six classes, where the x-axis is the manual count made by an expert and the y-axis is the automatic count generated by our mobile device. In these results, we observe a relatively large overcounting for Alternaria and undercounting for Aspergillus in **Figure 3.12(e,f)**, as also seen by their larger root mean square error (RMSE). This may be related to the fact that (1) the mold spores are smaller and therefore relatively more challenging to classify using the current resolution of our system, and (2) the mold spores tend to coagulate due to moisture, which may confuse the CNN model when they are present in the same ROI [24]. These results might be further improved using per-pixel semantic segmentation [196] instead of performing classification with a fixed window size.

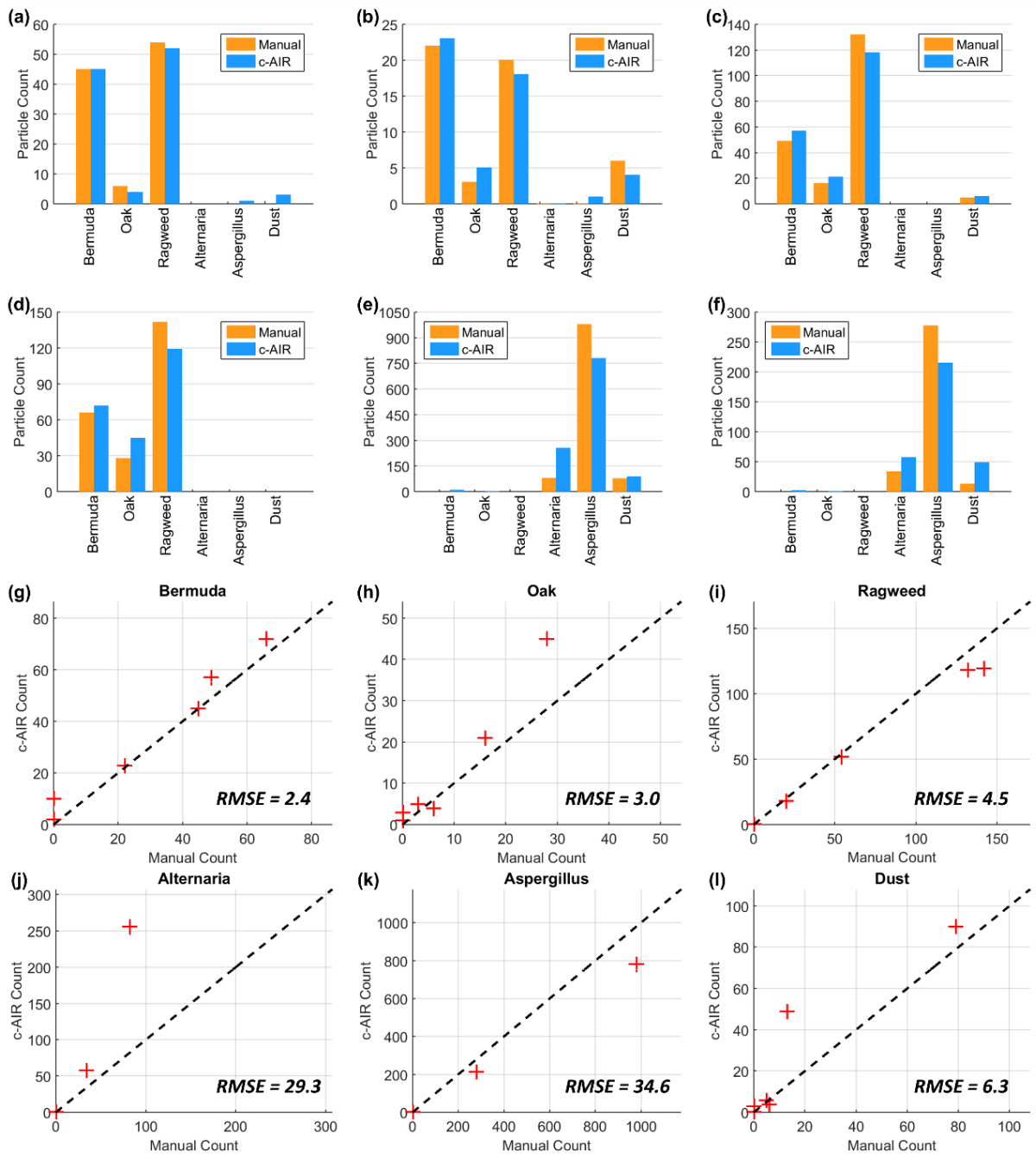

Figure 3.12 Bio-aerosol mixture experiments. (a-f) Deep-learning based automatic bio-aerosol detection and counting using our mobile device for six different experiments with varying bio-aerosol concentrations, and their comparisons against manual counting performed by a microbiologist under a benchtop scanning microscope with 40× magnification are shown. (g-l) Quantification of our counting accuracy for different types of aerosols. The dashed line refers to y = x. Root mean square error (RMSE) is also shown in each sub-figure. Reprinted from Ref. [24]

Field sensing of oak tree pollens

In the Spring of 2018, I used c-Air to measure bio-aerosols in air close to a line of four oak trees (Quercus Virginiana) at the University of California, Los Angeles campus [319]. A three-minute air sample is taken close to these trees at a pumping rate of 13 L / min, as illustrated in **Figure 3.13(a)**.

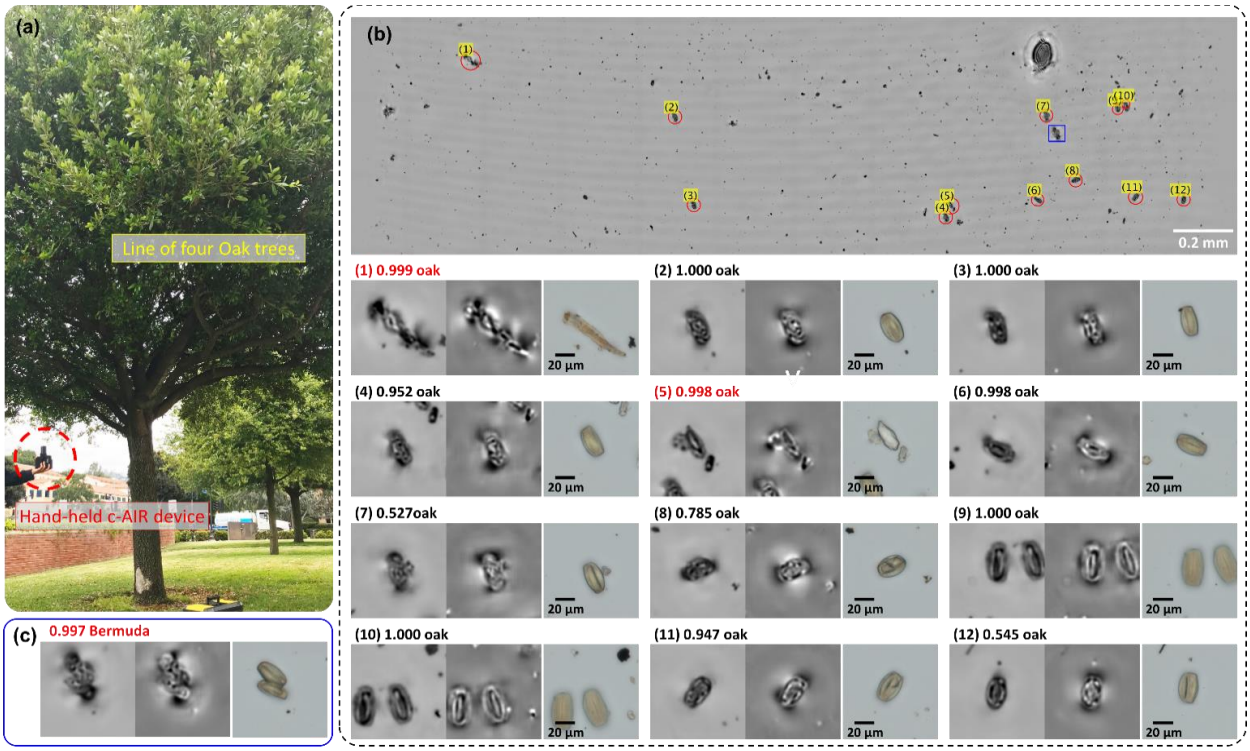

Figure 3.13 Sensing of oak tree pollens in the field. (a) Field testing of c-Air device is performed under a line of four oak trees in Los Angeles (Spring of 2018). (b) Full-FOV reconstruction of the captured aerosol samples is shown, where oak pollen bio-aerosols that are detected by deep learning-based classification algorithm are marked by red circles. The zoomed-in images of these detected particles, with real (left) and imaginary (right) images, reconstructed also using a deep neural network, are shown in (1) – (12). A comparison image captured later using a benchtop microscope under 40× magnification is also shown for each region. Softmax classification scores for each captured aerosol are also shown on top of each ROI. The two misclassification cases, (1) and (5), are marked in red. (c) A cluster of oak particles is misclassified as Bermuda pollen. Its location is highlighted by a blue square in (b). Reprinted from Ref. [24]

The whole FOV reconstruction of this sample is shown in **Figure 3.13(b)**, which also highlights different ROIs corresponding to the oak tree pollens automatically detected by the deep learning-based algorithm. In these 12 ROIs, there are two false positive detections (**Figure 3.13(b-1)** and **Figure 3.13(b-5)**), which are actually plant fragments that have elongated shapes. This problem might be addressed in the future by including such plant fragment images in our training dataset as an additional label. Next, I examined the entire FOV to screen for the false negative detections of oak tree pollens. Of all the detected bio-aerosols, we see that the CNN missed one cluster of oak tree pollens within the FOV, as marked by a blue rectangle in **Figure 3.13(b)** and shown in **Figure 3.13(c)**, which is classified as Bermuda with a high score. From **Figure 3.13(c)**, we see that this image contains two oak tree pollens clustered together, and since the training dataset only included isolated oak tree pollens it was misclassified as a Bermuda grass pollen, which is generally larger in size and rounder in shape than an oak tree pollen. Although the occurrence of clustered pollens is relatively rare, these types of misclassifications may be reduced by including clusters of pollen examples in the training dataset, or using per-pixel semantic segmentation instead of a classification CNN.

3.2.3 Conclusion

In summary, this mobile bio-aerosol sensing platform is hand-held, cost-effective and accurate. I envision that it can be used to build a wide-coverage automated bio-aerosol monitoring network in a cost-effective and scalable manner, which can rapidly provide accurate response for spatio-temporal mapping of bio-aerosol concentrations. The device is controlled wirelessly and can potentially be carried by unmanned vehicles such as drones to access bio-aerosol monitoring sites that may be dangerous for human inspectors.

3.2.4 Methods

Computational-imaging-based bio-aerosol monitoring

The c-Air device design is similar to the one described in section 3.1.4 and in **Figure 3.1**, with minor modifications. In this case, an infrared vertical-cavity surface-emitting laser (VCSEL) diode (OPV300, TT Electronics, $\lambda_p = 850\,nm$) illuminates the collected aerosols from above, casting an in-line hologram of the samples, which is recorded by a newer CMOS image sensor chip (Sony IMX219PQ, pixel pitch 1.12 μm). These in-line holograms are sent to a remote server where the aerosol images are analyzed automatically. To avoid secondary light sources from the reflection and refraction of the transparent impactor nozzle, a 3D-printed black cover is used to tightly cover the impactor surface. A driver chip (TLC5941NT, Texas Instruments) controls the current of our illumination VCSEL at its threshold (3 mA), which provides adequate coherence without introducing speckle noise. 850 nm illumination wavelength is specifically chosen to use all of the four Bayer channels on the color CMOS imager, since all the four Bayer channels have equal transmission at this wavelength, making it function like a monochrome sensor for our holographic imaging purposes. Benefited from this, as well as higher coherence of the laser diode, a better spatial resolution is achieved. The entire mobile device is powered by a Lithium polymer (Li-po) battery (Turnigy Nano-tech 1000mAh 4S 45~90C Li-po pack) and controlled by a newer embedded single board computer (Raspberry Pi Zero W).

Simultaneous autofocusing and phase recovery of bio-aerosols using deep learning

To simultaneously perform digital autofocusing and phase recovery for each individual aerosol, a CNN-based method is used [129], built using Tensorflow [110]. This CNN is trained using HIDEF approach (section 2.3), with up to a defocus distance of ± 100 μm. Due to limited graphical memory of our computer, the full FOV back-propagated image (9840 × 3069 × 2) cannot be

processed directly; it is therefore divided into 12-by-5 patches of 1024-by-1024 pixels with a spatial overlap of 100-pixels between images. Each individual patch is processed in sequence and the results are combined after this reconstruction step to reveal the bio-aerosols captured within the entire FOV. Each patch takes ~ 0.4 s to process, totaling ~ 25 s for the entire FOV.

Aerosol detection algorithm

A multi-scale spot detection algorithm similar to Ref. [320] is developed to detect and extract each aerosol ROI. The algorithm takes six levels of high pass filtering of the complex amplitude image per ROI, obtained by the difference of the original image and the blurred images filtered by six different kernels. These high-passed images are per-pixel multiplied with each other to obtain a correlation image. A binary mask is then obtained by thresholding this correlation image with three-times the standard deviation added to the mean of the correlation image. This binary mask is dilated by 11 pixels, and the connected components are used to estimate a circle with the centroid and radius of each one of the detected spots, which marks the location and rough size of each detected bio-aerosol. To avoid multiple detections of the same aerosol, a non-maximum suppression criterion is applied, where if an estimated circle has more than 10% of overlapping area with another circle, only the bigger one is considered/counted. The resulting centroids are cropped into 256×256 pixel ROIs, which are then fed into the classification CNN, detailed in the next sub-section. This algorithm takes < 5 s for the whole FOV, and achieves better performance compared to conventional circle detection algorithms such as circular Hough transform [321], achieving 98.4% detection precision and 92.5% detection recall [24].

Deep learning-based classification of bio-aerosols

The classification CNN architecture is shown in the zoom-in part of **Figure 3.10(b)**, which is inspired by ResNet [103]. The network contains five residual blocks, where each block maps the input tensor x_k into output tensor x_{k+1}, for a given block k ($k = 1,2,3,4,5$), i.e.,

$$\mathrm{x}'_{\mathrm{k}} = \mathrm{x}_{\mathrm{k}} + \mathrm{ReLU}\left[\mathrm{CONV}_{\mathrm{k}_1}\left\{\mathrm{ReLU}\left[\mathrm{CONV}_{\mathrm{k}_1}\{\mathrm{x}_{\mathrm{k}}\}\right]\right\}\right] \tag{3.13}$$

$$\mathrm{x}_{\mathrm{k+1}} = \mathrm{MAX}\left(\mathrm{x}'_{\mathrm{k}} + \mathrm{ReLU}\left[\mathrm{CONV}_{\mathrm{k}_2}\left\{\mathrm{ReLU}\left[\mathrm{CONV}_{\mathrm{k}_1}\{\mathrm{x}'_{\mathrm{k}}\}\right]\right\}\right]\right) \tag{3.14}$$

where ReLU stands for rectified linear unit operation, CONV stands for the convolution operator (including the bias terms), and MAX stands for the max-pooling operator. The subscript k_1 and k_2 denote the number of channels in the corresponding convolution layer, where k_1 equals to the number of input channels and k_2 expands the number of channels twice, i.e. $k_1 = 16, 32, 64, 128, 256$ and $k_2 = 32, 64, 128, 256, 512$ for each residual block ($k = 1,2,3,4,5$). Zero padding is used on the tensor x'_{k} to compensate the mismatch between the number of input and output channels. All the convolutional layers use a convolutional kernel of 3×3 pixels, a stride of one pixel, and a replicate-padding of one pixel. All the max-pooling layers use a kernel of two pixels, and a stride of two pixels, which reduces the width and height of the image by half. Following the residual blocks, an average pooling layer reduces the width and height of the tensor to one, which is followed by a fully-connected (FC) layer of size 512×512. Dropout with 0.5 probability is used on this FC layer to increase performance and prevent overfitting. Another fully connected layer of size 512×6 maps the 512 channels to 6 class scores for final determination of the class of each imaged bio-aerosol.

During training, the network minimizes the soft-max cross-entropy loss between the true label and the output scores:

$$\mathrm{L} = \sum_{\mathrm{i}} \left[-\log\left(\frac{\mathrm{e}^{\mathrm{f}_{\mathrm{y}_\mathrm{i}}(\mathrm{x}_\mathrm{i})}}{\sum_{\mathrm{j}} \mathrm{e}^{\mathrm{f}_{\mathrm{j}}(\mathrm{x}_\mathrm{i})}} \right) \right] \tag{3.15}$$

where $f_j(x_i)$ is the class score for the class j given input data x_i, and y_i is the corresponding true class for x_i. The dataset contains ~1,500 individual 256×256 pixel ROIs for each of the six classes, totaling ~10,000 images. 70% of the data for each class is randomly selected for training, and remaining images are equally divided to validation and testing sets. The training takes ~2 h for 200 epochs. The best trained model is selected to be the one that gives lowest soft-max loss for the validation set within 200 training epochs. The testing takes < 0.02 s for each 256×256 pixel ROI. For a typical FOV with e.g., ~500 particles, this step is completed in ~10 s.

Chapter 4 Conclusion

Computational imaging, which relies on computational algorithms to form the image, overcomes various limitations of conventional hardware-centered imaging paradigm and enabled many breakthroughs in optical microscopy. Recently, we are experiencing a renaissance in computational microscopy, enabled by novel statistical tools which stems from deep neural networks and data-driven machine learning. Compared to analytical and iterative solutions (Chapter 1), emerging deep neural networks learns *content-aware priors* that can be considered as a *task-specific regularizer* to the inverse problems, which offers unprecedented performance in various image reconstruction tasks including image super resolution, phase recovery, and transformation between different microscope modalities and contrast mechanisms.

In summary of Chapter 2, my contributions to deep learning based computational imaging include: **(1)** *I* ***introduced HIDEF approach that solves digital holographic phase recovery and auto-focusing simultaneously***. The HIDEF approach is non-iterative and quickly generates an extended DOF phase-recovered reconstruction from a single hologram recording, which fundamentally reduces the required measurement from 4-8 to a single hologram, and improves the algorithm complexity of digital holographic image reconstruction from *O(nm)* to *O(1)*. HIDEF has since been widely used for rapid holographic reconstruction of planktons, pollens and smaller molecules. **(2)** ***I designed "Bright-field holography" approach that bridges the contrast gap between coherent and incoherent imaging modalities using deep CNN.*** Coherent imaging such as digital holography enables 3D imaging of objects from a snapshot hologram, benefiting from the wave-propagation framework. However, the reconstructed image contrast is inferior to incoherent modalities like bright field microscopy due to various sources of coherent

noises. Bright-field holography uses a GAN to transform a back-propagated hologram into the equivalent image scanned by a high-NA bright field microscope, fusing the advantages of coherent and incoherent imaging into the same framework. **(3)** ***I invented Deep-Z for 3D virtual refocusing of fluorescence images onto user defined surfaces.*** Fluorescence microscopy generally lacks the propagation framework and requires axial scanning to image 3D samples. In Deep-Z, a neural network learns the defocus behavior of fluorescence and enables such refocusing framework, providing 20-fold DOF extension without axial scanning, additional hardware, or a trade-off of imaging resolution or speed.

Deep learning has also enabled computational environmental sensing, including air quality monitoring. In summary of Chapter 3, my contributions here include: **(4)** ***I made c-Air, which is a portable computational microscope, powered by machine learning, for aerosol monitoring.*** c-Air uses lens-less digital holographic imaging geometry and provides high-throughput, (screens 6.5 liters of air in 30 seconds), cost-effective and portable sensing of particulate matters, with a sizing accuracy of 93%. Using c-Air, I reported intriguing results for PM monitoring during forest fires and near major airports. **(5)** ***Based on c-Air, I provided a novel framework for label-free bio-aerosol sensing using deep learning***. Using CNNs, I achieved > 94% accuracy in label-free sensing and classification of some of the most common bio-aerosols, automatically using the portable c-Air device.

Microscopy is an ideal platform for deep learning as the instruments are high-precision and the results are highly repeatable. In the long run, I believe deep learning will play an increasingly more important role in computational optical microscopy and environmental sensing, which will be driven by the design of novel learning algorithms, network architectures, improvements in parallel computing resources, as well as new applications. Specifically, deep learning-based

image reconstruction techniques, such as *HIDEF*, demonstrates the power of data-driven approach, and will continue to find wider applications. *Bright-field holography*, as well as other cross-modality/contrast transformations enabled by deep learning, will act like a catalyst that accelerate the wide adoption of holography and other coherent imaging techniques in biomedical applications. *Deep-Z* provides a unique framework for incoherent image digital refocusing, which will be especially useful for imaging fast events of live samples with significantly reduced photo-damage. The *c-Air device*, with its mobility and cost-effectiveness, and powered by deep learning, will find several commercial applications in indoor and outdoor air quality monitoring.

References

1. Bardell, D. The Biologists' Forum: The invention of the microscope. *BIOS* **75**, 78–84 (2004).
2. McLeod, E. & Ozcan, A. Unconventional methods of imaging: computational microscopy and compact implementations. *Rep. Prog. Phys.* **79**, 076001 (2016).
3. Ozcan, A. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab. Chip* **14**, 3187–3194 (2014).
4. Greenbaum, A. *et al.* Wide-field computational imaging of pathology slides using lens-free on-chip microscopy. *Sci. Transl. Med.* **6**, 267ra175-267ra175 (2014).
5. Brooker, G. *et al.* In-line FINCH super resolution digital holographic fluorescence microscopy using a high efficiency transmission liquid crystal GRIN lens. *Opt. Lett.* **38**, 5264–5267 (2013).
6. Bianco, V. *et al.* Optofluidic holographic microscopy with custom field of view (FoV) using a linear array detector. *Lab. Chip* **15**, 2117–2124 (2015).
7. Laporte, G. P., Stasio, N., Sheppard, C. J. & Psaltis, D. Resolution enhancement in nonlinear scanning microscopy through post-detection digital computation. *Optica* **1**, 455–460 (2014).
8. McLeod, E. *et al.* High-throughput and label-free single nanoparticle sizing based on time-resolved on-chip microscopy. *ACS Nano* **9**, 3265–3273 (2015).
9. Rivenson, Y., Stern, A. & Javidi, B. Improved depth resolution by single-exposure in-line compressive holography. *Appl. Opt.* **52**, A223–A231 (2013).
10. Rosen, J. & Brooker, G. Non-scanning motionless fluorescence three-dimensional holographic microscopy. *Nat. Photonics* **2**, 190–195 (2008).

11. Stern, A. & Javidi, B. Theoretical analysis of three-dimensional imaging and recognition of micro-organisms with a single-exposure on-line holographic microscope. *J. Opt. Soc. Am. A* **24**, 163–168 (2007).

12. Barsi, C., Wan, W. & Fleischer, J. W. Imaging through nonlinear media using digital holography. *Nat. Photonics* **3**, 211–215 (2009).

13. Ferraro, P. *et al.* Full color 3-D imaging by digital holography and removal of chromatic aberrations. *J. Disp. Technol.* **4**, 97–100 (2008).

14. Kim, T. *et al.* White-light diffraction tomography of unlabelled live cells. *Nat. Photonics* **8**, 256–263 (2014).

15. Greenbaum, A. *et al.* Increased space-bandwidth product in pixel super-resolved lensfree on-chip microscopy. *Sci. Rep.* **3**, (2013).

16. Greenbaum, A. *et al.* Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy. *Nat. Methods* **9**, 889–895 (2012).

17. Mudanyali, O. *et al.* Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications. *Lab. Chip* **10**, 1417–1428 (2010).

18. Zhang, Y. *et al.* Wide-field imaging of birefringent synovial fluid crystals using lens-free polarized microscopy for gout diagnosis. *Sci. Rep.* **6**, (2016).

19. Zhang, Y., Greenbaum, A., Luo, W. & Ozcan, A. Wide-field pathology imaging using on-chip microscopy. *Virchows Arch.* **467**, 3–7 (2015).

20. O. Isikman, S. *et al.* Color and monochrome lensless on-chip imaging of Caenorhabditis elegans over a wide field-of-view. *Lab. Chip* **10**, 1109–1112 (2010).

21. Isikman, S. O., Bishara, W., Zhu, H. & Ozcan, A. Optofluidic tomography on a chip. *Appl. Phys. Lett.* **98**, 161109 (2011).

22. Mudanyali, O., Oztoprak, C., Tseng, D., Erlinger, A. & Ozcan, A. Detection of waterborne parasites using field-portable and cost-effective lensfree microscopy. *Lab Chip* **10**, 2419–2423 (2010).

23. Wu, Y.-C. *et al.* Air quality monitoring using mobile microscopy and machine learning. *Light Sci. Appl.* **6**, e17046–e17046 (2017).

24. Wu, Y. *et al.* Label-Free Bioaerosol Sensing Using Mobile Microscopy and Deep Learning. *ACS Photonics* **5**, 4617–4627 (2018).

25. Feizi, A. *et al.* 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).

26. Su, T.-W., Erlinger, A., Tseng, D. & Ozcan, A. Compact and light-weight automated semen analysis platform using lensfree on-chip microscopy. *Anal. Chem.* **82**, 8307–8312 (2010).

27. Su, T.-W., Xue, L. & Ozcan, A. High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories. *Proc. Natl. Acad. Sci.* **109**, 16018–16022 (2012).

28. Su, T.-W. *et al.* Sperm trajectories form chiral ribbons. *Sci. Rep.* **3**, (2013).

29. Su, T.-W., Choi, I., Feng, J., Huang, K. & Ozcan, A. High-throughput analysis of horse sperms' 3D swimming patterns using computational on-chip imaging. *Anim. Reprod. Sci.* **169**, 45–55 (2016).

30. Luo, W. *et al.* High throughput on-chip analysis of high-energy charged particle tracks using lensfree imaging. *Appl. Phys. Lett.* **106**, 151107 (2015).

31. Huang, K.-W., Su, T.-W., Ozcan, A. & Chiou, P.-Y. Optoelectronic tweezers integrated with lensfree holographic microscopy for wide-field interactive cell and particle manipulation on a chip. *Lab. Chip* **13**, 2278–2284 (2013).

32. Rivenson, Y. *et al.* Sparsity-based multi-height phase recovery in holographic microscopy. *Sci. Rep.* **6**, 37862 (2016).

33. Wu, Y., Zhang, Y., Luo, W. & Ozcan, A. Demosaiced pixel super-resolution for multiplexed holographic color imaging. *Sci. Rep.* **6**, 28601 (2016).

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

35. Goodman, J. W. *Introduction to Fourier optics, 3rd Edition*. (Roberts and Company Publishers, 2005).

36. Seo, S., Su, T.-W., Tseng, D. K., Erlinger, A. & Ozcan, A. Lensfree holographic imaging for on-chip cytometry and diagnostics. *Lab. Chip* **9**, 777–787 (2009).

37. Seo, S. *et al.* High-Throughput Lens-Free Blood Analysis on a Chip. *Anal. Chem.* **82**, 4621–4627 (2010).

38. Luo, W., Greenbaum, A., Zhang, Y. & Ozcan, A. Synthetic aperture-based on-chip microscopy. *Light Sci. Appl.* **4**, e261 (2015).

39. Tseng, D. *et al.* Lensfree microscopy on a cellphone. *Lab. Chip* **10**, 1787–1792 (2010).

40. Greenbaum, A., Akbari, N., Feizi, A., Luo, W. & Ozcan, A. Field-portable pixel super-resolution colour microscope. *PloS One* **8**, e76475 (2013).

41. Isikman, S. O. *et al.* Lens-free optical tomographic microscope with a large imaging volume on a chip. *Proc. Natl. Acad. Sci.* **108**, 7296–7301 (2011).

42. Born, M. & Wolf, E. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. (CUP Archive, 2000).

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

44. Farsiu, S., Robinson, M. D., Elad, M. & Milanfar, P. Fast and robust multiframe super resolution. *IEEE Trans. Image Process.* **13**, 1327–1344 (2004).

45. Hardie, R. C., Barnard, K. J., Bognar, J. G., Armstrong, E. E. & Watson, E. A. High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Opt. Eng.* **37**, 247–260 (1998).

46. Bishara, W., Su, T.-W., Coskun, A. F. & Ozcan, A. Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution. *Opt. Express* **18**, 11181–11191 (2010).

47. Bishara, W., Zhu, H. & Ozcan, A. Holographic opto-fluidic microscopy. *Opt. Express* **18**, 27499–27510 (2010).

48. Bishara, W. *et al.* Holographic pixel super-resolution in portable lensless on-chip microscopy using a fiber-optic array. *Lab. Chip* **11**, 1276–1279 (2011).

49. Farsiu, S., Elad, M. & Milanfar, P. Multiframe demosaicing and super-resolution of color images. *IEEE Trans. Image Process.* **15**, 141–159 (2006).

50. Coskun, A. F., Sencan, I., Su, T.-W. & Ozcan, A. Lensless wide-field fluorescent imaging on a chip using compressive decoding of sparse objects. *Opt. Express* **18**, 10510–10523 (2010).

51. Chambolle, A. An Algorithm for Total Variation Minimization and Applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004).

52. Boyd, S. & Vandenberghe, L. *Convex optimization*. (Cambridge university press, 2004).

53. Luo, W., Zhang, Y., Feizi, A., Gorocs, Z. & Ozcan, A. Pixel super-resolution using wavelength scanning. *Light Sci. Appl.* **5**, e16058 (2016).

54. Luo, W., Zhang, Y., Göröcs, Z., Feizi, A. & Ozcan, A. Propagation phasor approach for holographic image reconstruction. *Sci. Rep.* **6**, 22738 (2016).

55. Wang, H. *et al.* Computational out-of-focus imaging increases the space–bandwidth product in lens-based coherent microscopy. *Optica* **3**, 1422–1429 (2016).

56. Kim, J., Lee, J. K. & Lee, K. M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. *ArXiv151104587 Cs* (2015).

57. Kim, J., Lee, J. K. & Lee, K. M. Deeply-Recursive Convolutional Network for Image Super-Resolution. *ArXiv151104491 Cs* (2015).

58. Dong, C., Loy, C. C., He, K. & Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 295–307 (2016).

59. Shi, W. *et al.* Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *ArXiv160905158 Cs Stat* (2016).

60. Rivenson, Y. *et al.* Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).

61. Oppenheim, A. V., Willsky, A. S. & Nawab, S. H. *Signals and systems*. vol. 2 (Prentice-Hall Englewood Cliffs, NJ, 1983).

62. Memmolo, P. *et al.* Automatic focusing in digital holography and its application to stretched holograms. *Opt. Lett.* **36**, 1945–1947 (2011).

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

64. Memmolo, P., Paturzo, M., Javidi, B., Netti, P. A. & Ferraro, P. Refocusing criterion via sparsity measurements in digital holography. *Opt. Lett.* **39**, 4719–4722 (2014).

65. Langehanenberg, P., Kemper, B., Dirksen, D. & Bally, G. von. Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging. *Appl. Opt.* **47**, D176–D182 (2008).

66. Dubois, F., Schockaert, C., Callens, N. & Yourassowsky, C. Focus plane detection criteria in digital holography microscopy by amplitude analysis. *Opt. Express* **14**, 5895–5908 (2006).

67. Liebling, M. & Unser, M. Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion. *JOSA A* **21**, 2424–2430 (2004).

68. Dubois, F., Mallahi, A. E., Dohet-Eraly, J. & Yourassowsky, C. Refocus criterion for both phase and amplitude objects in digital holographic microscopy. *Opt. Lett.* **39**, 4286–4289 (2014).

69. Lyu, M., Yuan, C., Li, D. & Situ, G. Fast autofocusing in digital holography using the magnitude differential. *Appl. Opt.* **56**, F152–F157 (2017).

70. Tamamitsu, M., Zhang, Y., Wang, H., Wu, Y. & Ozcan, A. Comparison of Gini index and Tamura coefficient for holographic autofocusing based on the edge sparsity of the complex optical wavefront. *ArXiv170808055 Phys.* (2017).

71. Fienup, J. Phase Retrieval Algorithms - a Comparison. *Appl. Opt.* **21**, 2758–2769 (1982).

72. Biener, G. *et al.* Combined reflection and transmission microscope for telemedicine applications in field settings. *Lab. Chip* **11**, 2738–2743 (2011).

73. Gerchber, R. W. & Saxton, W. O. Practical Algorithm for Determination of Phase from Image and Diffraction Plane Pictures. *Optik* **35**, 237- (1972).

74. Elser, V. Phase retrieval by iterated projections. *J. Opt. Soc. Am. A* **20**, 40 (2003).

75. Allen, L. J. & Oxley, M. P. Phase retrieval from series of images obtained by defocus variation. *Opt. Commun.* **199**, 65–75 (2001).

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

77. Fienup, J. & Wackerman, C. Phase-Retrieval Stagnation Problems and Solutions. *J. Opt. Soc. Am. -Opt. Image Sci. Vis.* **3**, 1897–1907 (1986).

78. Gureyev, T. E. & Nugent, K. A. Phase retrieval with the transport-of-intensity equation. II. Orthogonal series solution for nonuniform illumination. *JOSA A* **13**, 1670–1682 (1996).

79. Waller, L., Tian, L. & Barbastathis, G. Transport of Intensity phase-amplitude imaging with higher order intensity derivatives. *Opt. Express* **18**, 12552–12561 (2010).

80. Weidling, J., Isikman, S. O., Greenbaum, A., Ozcan, A. & Botvinick, E. Lens-free computational imaging of capillary morphogenesis within three-dimensional substrates. *J. Biomed. Opt.* **17**, 126018–126018 (2012).

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

82. Waller, L., Tsang, M., Ponda, S., Yang, S. Y. & Barbastathis, G. Phase and amplitude imaging from noisy images by Kalman filtering. *Opt. Express* **19**, 2805–2814 (2011).

83. Jingshan, Z., Tian, L., Dauwels, J. & Waller, L. Partially coherent phase imaging with simultaneous source recovery. *Biomed. Opt. Express* **6**, 257–265 (2015).

84. Jingshan, Z., Dauwels, J., Vázquez, M. A. & Waller, L. Sparse ACEKF for phase reconstruction. *Opt. Express* **21**, 18125–18137 (2013).

85. Rudin, L. I., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. Nonlinear Phenom.* **60**, 259–268 (1992).

86. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B-Methodol.* **58**, 267–288 (1996).

87. Ballard, Z. S. *et al.* Computational Sensing Using Low-Cost and Mobile Plasmonic Readers Designed by Machine Learning. *ACS Nano* **11**, 2266–2274 (2017).

88. Rivenson, Y., Stern, A. & Javidi, B. Overview of compressive sensing techniques applied in holography [Invited]. *Appl. Opt.* **52**, A423–A432 (2013).

89. Latychevskaia, T. & Fink, H.-W. Solution to the Twin Image Problem in Holography. *Phys. Rev. Lett.* **98**, 233901 (2007).

90. Rivenson, Y., Stern, A. & Rosen, J. Compressive multiple view projection incoherent holography. *Opt. Express* **19**, 6109–6118 (2011).

91. Brady, D. J., Choi, K., Marks, D. L., Horisaki, R. & Lim, S. Compressive holography. *Opt. Express* **17**, 13040–13049 (2009).

92. Sencan, I., Coskun, A. F., Sikora, U. & Ozcan, A. Spectral demultiplexing in holographic and fluorescent on-chip microscopy. *Sci. Rep.* **4**, (2014).

93. Coskun, A. F., Sencan, I., Su, T.-W. & Ozcan, A. Wide-field lensless fluorescent microscopy using a tapered fiber-optic faceplate on a chip. *Analyst* **136**, 3512–3518 (2011).

94. Coskun, A. F., Sencan, I., Su, T.-W. & Ozcan, A. Lensfree fluorescent on-chip imaging of transgenic Caenorhabditis elegans over an ultra-wide field-of-view. *PLoS One* **6**, e15955–e15955 (2011).

95. Zhu, H., Mavandadi, S., Coskun, A. F., Yaglidere, O. & Ozcan, A. Optofluidic fluorescent imaging cytometry on a cell phone. *Anal. Chem.* **83**, 6641–6647 (2011).

96. Greenbaum, A., Feizi, A., Akbari, N. & Ozcan, A. Wide-field computational color imaging using pixel super-resolved on-chip microscopy. *Opt. Express* **21**, 12469–12483 (2013).

97. Göröcs, Z., Kiss, M., Tóth, V., Orzó, L. & Tokés, S. Multicolor digital holographic microscope (DHM) for biological purposes. in *BiOS* (eds. Farkas, D. L., Nicolau, D. V. & Leif, R. C.) 75681P-75681P–10 (International Society for Optics and Photonics, 2010). doi:10.1117/12.841962.

98. Göröcs, Z., Orzó, L., Kiss, M., Tóth, V. & Tőkés, S. In-line color digital holographic microscope for water quality measurements. in *Laser Applications in Life Sciences 2010* (eds. Kinnunen, M. & Myllylä, R.) 737614-737614–10 (International Society for Optics and Photonics, 2010). doi:10.1117/12.871098.

99. Kiss, M. Z. *et al.* Special multicolor illumination and numerical tilt correction in volumetric digital holographic microscopy. *Opt. Express* **22**, 7559–7573 (2014).

100. Zhang, Y., Wu, Y., Zhang, Y. & Ozcan, A. Color calibration and fusion of lens-free and mobile-phone microscopy images for high-resolution and accurate color reproduction. *Sci. Rep.* **6**, 27811 (2016).

101. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems 25* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 1097–1105 (Curran Associates, Inc., 2012).

102. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs* (2014).

103. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016). doi:10.1109/CVPR.2016.90.

104. Gonzalez, R. C., Woods, R. E. & Eddins, S. L. Digital image processing using MATLAB. (2004).

105. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943).

106. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533 (1986).

107. LeCun, Y. *et al.* Handwritten digit recognition with a back-propagation network. in *Advances in neural information processing systems* 396–404 (1990).

108. Bengio, Y., Simard, P. & Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **5**, 157–166 (1994).

109. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).

110. Abadi, M. *et al.* TensorFlow: A System for Large-Scale Machine Learning. in *OSDI* vol. 16 265–283 (2016).

111. Paszke, A. *et al.* Automatic differentiation in PyTorch. in (2017).

112. Jia, Y. *et al.* Caffe: Convolutional Architecture for Fast Feature Embedding. *ArXiv14085093 Cs* (2014).

113. The Theano Development Team *et al.* Theano: A Python framework for fast computation of mathematical expressions. *ArXiv160502688 Cs* (2016).

114. Szegedy, C. *et al.* Going Deeper With Convolutions. in 1–9 (2015).

115. Young, T., Hazarika, D., Poria, S. & Cambria, E. Recent Trends in Deep Learning Based Natural Language Processing [Review Article]. *IEEE Comput. Intell. Mag.* **13**, 55–75 (2018).

116. Silver, D. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

117. Eslami, S. M. A. *et al.* Neural scene representation and rendering. *Science* **360**, 1204–1210 (2018).

118. Christiansen, E. M. *et al.* In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* **173**, 792-803.e19 (2018).

119. Cox, I. J. & Sheppard, C. J. R. Information capacity and resolution in an optical system. *JOSA A* **3**, 1152–1158 (1986).

120. Katznelson, Y. *An Introduction to Harmonic Analysis*. (Cambridge University Press, 2004).

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

122. Rivenson, Y. *et al.* Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).

123. Wu, Y. *et al.* Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nat. Methods* **accepted**, (2019).

124. Wang, H. *et al.* Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).

125. Liu, T. *et al.* Deep learning-based super-resolution in coherent imaging systems. *Sci. Rep.* **9**, 3926 (2019).

126. de Haan, K., Ballard, Z. S., Rivenson, Y., Wu, Y. & Ozcan, A. Resolution enhancement in scanning electron microscopy using deep learning. *arXiv:1901.11094* (2019).

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

128. Rivenson, Y., Zhang, Y., Günaydın, H., Teng, D. & Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).

129. Wu, Y. *et al.* Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* **5**, 704–710 (2018).

130. Goy, A., Arthur, K., Li, S. & Barbastathis, G. Low Photon Count Phase Retrieval Using Deep Learning. *Phys. Rev. Lett.* **121**, 243902 (2018).

131. Rivenson, Y., Wu, Y. & Ozcan, A. Deep learning in holography and coherent imaging. *Light Sci. Appl.* **8**, 1–8 (2019).

132. Jeon, W., Jeong, W., Son, K. & Yang, H. Speckle noise reduction for digital holographic images using multi-scale convolutional neural networks. *Opt. Lett.* **43**, 4240–4243 (2018).

133. Li, S., Deng, M., Lee, J., Sinha, A. & Barbastathis, G. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).

134. Li, Y., Xue, Y. & Tian, L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).

135. Borhani, N., Kakkava, E., Moser, C. & Psaltis, D. Learning to see through multimode fibers. *Optica* **5**, 960–966 (2018).

136. Kamilov, U. S. *et al.* Learning approach to optical tomography. *Optica* **2**, 517–522 (2015).

137. Choi, G. *et al.* Cycle-consistent deep learning approach to coherent noise reduction in optical diffraction tomography. *Opt. Express* **27**, 4927–4943 (2019).

138. Goy, A. *et al.* High-Resolution Limited-Angle Phase Tomography of Dense Layered Objects Using Deep Neural Networks. *ArXiv181207380 Phys.* (2018).

139. Wu, Y. *et al.* Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl.* **8**, 25 (2019).

140. Rivenson, Y. *et al.* Deep Learning Enhanced Mobile-Phone Microscopy. *ACS Photonics* **5**, 2354–2364 (2018).

141. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).

142. Liu, Y. *et al.* Detecting Cancer Metastases on Gigapixel Pathology Images. *CoRR* **abs/1703.02442**, (2017).

143. Rivenson, Y. *et al.* Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* 1 (2019) doi:10.1038/s41551-019-0362-y.

144. Borhani, N., Bower, A. J., Boppart, S. A. & Psaltis, D. Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy. *Biomed. Opt. Express* **10**, 1339–1350 (2019).

145. Vakoc, B. J. *et al.* Three-dimensional microscopy of the tumor microenvironment in vivo using optical frequency domain imaging. *Nat. Med.* **15**, 1219–1223 (2009).

146. Dalimier, E. & Salomon, D. Full-Field Optical Coherence Tomography: A New Technology for 3D High-Resolution Skin Imaging. *Dermatology* **224**, 84–92 (2012).

147. Rivenson, Y. *et al.* PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci. Appl.* **8**, 23 (2019).

148. Lu, F.-K. *et al.* Label-free DNA imaging in vivo with stimulated Raman scattering microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11624–11629 (2015).

149. Orringer, D. A. *et al.* Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nat. Biomed. Eng.* **1**, (2017).

150. Tao, Y. K. *et al.* Assessment of breast pathologies using nonlinear microscopy. *Proc. Natl. Acad. Sci.* **111**, 15304–15309 (2014).

151. Fereidouni, F. *et al.* Microscopy with ultraviolet surface excitation for rapid slide-free histology. *Nat. Biomed. Eng.* **1**, 957–966 (2017).

152. Glaser, A. K. *et al.* Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens. *Nat. Biomed. Eng.* **1**, (2017).

153. Christiansen, E. M. *et al.* In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* **173**, 792–803 (2018).

154. Dardikman, G. & Shaked, N. T. Phase Unwrapping Using Residual Neural Networks. in *Imaging and Applied Optics 2018 (3D, AO, AIO, COSI, DH, IS, LACSEA, LS&C, MATH, pcAOP) (2018), paper CW3B.5* CW3B.5 (Optical Society of America, 2018). doi:10.1364/COSI.2018.CW3B.5.

155. Spoorthi, G. E., Gorthi, S. & Gorthi, R. K. S. S. PhaseNet: A Deep Convolutional Neural Network for Two-Dimensional Phase Unwrapping. *IEEE Signal Process. Lett.* **26**, 54–58 (2019).

156. Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning Hierarchical Features for Scene Labeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1915–1929 (2013).

157. Nair, V. & Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. in *Proceedings of the 27th International Conference on Machine Learning* 807–814 (Omnipress, 2010).

158. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).

159. Pedamonti, D. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. *arXiv:1804.02763* (2018).

160. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput* **9**, 1735–1780 (1997).

161. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 234–241 (Springer International Publishing, 2015).

162. Bergstra, J. & Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012).

163. Neyshabur, B., Bhojanapalli, S., Mcallester, D. & Srebro, N. Exploring Generalization in Deep Learning. in *Advances in Neural Information Processing Systems 30* (eds. Guyon, I. et al.) 5947–5956 (Curran Associates, Inc., 2017).

164. Luo, W., Li, Y., Urtasun, R. & Zemel, R. Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems 29* (eds. Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. & Garnett, R.) 4898–4906 (Curran Associates, Inc., 2016).

165. Stanley, K. O. & Miikkulainen, R. Evolving Neural Networks Through Augmenting Topologies. *Evol Comput* **10**, 99–127 (2002).

166. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. in *Advances in Neural Information Processing Systems (NIPS)* (eds. Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q.) 2951–2959 (Curran Associates, Inc., 2012).

167. Zoph, B. & Le, Q. V. Neural Architecture Search with Reinforcement Learning. in *Proceedings of the International Conference on Learning Representations (ICLR)* (2017).

168. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67 (2019).

169. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495 (2017).

170. Mirza, M. & Osindero, S. Conditional Generative Adversarial Nets. *ArXiv14111784 Cs Stat* (2014).

171. Goshtasby, A. A. *2DD and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. (Wiley-Interscience, 2005).

172. Sokooti, H. *et al.* Nonrigid Image Registration Using Multi-scale 3D Convolutional Neural Networks. in *Medical Image Computing and Computer Assisted Intervention (MICCAI)* (eds. Descoteaux, M. et al.) 232–239 (Springer International Publishing, 2017).

173. Li, H. & Fan, Y. Non-rigid image registration using self-supervised fully convolutional networks without training data. in *IEEE 15th International Symposium on Biomedical Imaging (ISBI)* 1075–1078 (2018). doi:10.1109/ISBI.2018.8363757.

174. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917 (2018).

175. Zhu, J., Park, T., Isola, P. & Efros, A. A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. in *IEEE International Conference on Computer Vision (ICCV)* 2242–2251 (2017). doi:10.1109/ICCV.2017.244.

176. Olson, M., Wyner, A. & Berk, R. Modern Neural Networks Generalize on Small Data Sets. in *Advances in Neural Information Processing Systems (NIPS)* 3619–3628 (Curran Associates, Inc., 2018).

177. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How Transferable Are Features in Deep Neural Networks? in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)* vol. 2 3320–3328 (MIT Press, 2014).

178. Raina, R., Battle, A., Lee, H., Packer, B. & Ng, A. Y. Self-taught learning: transfer learning from unlabeled data. in 759–766 (ACM, 2007). doi:10.1145/1273496.1273592.

179. Sun, Y., Wang, X. & Tang, X. Deep Learning Face Representation from Predicting 10,000 Classes. in *2014 IEEE Conference on Computer Vision and Pattern Recognition* 1891–1898 (2014). doi:10.1109/CVPR.2014.244.

180. Perez, L. & Wang, J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv:1712.04621* (2017).

181. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

182. Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) no-reference image quality score - MATLAB brisque. https://www.mathworks.com/help/images/ref/brisque.html#d120e8865.

183. Xu, W., Jericho, M. H., Meinertzhagen, I. A. & Kreuzer, H. J. Digital in-line holography for biological applications. *Proc. Natl. Acad. Sci.* **98**, 11301–11305 (2001).

184. Marquet, P. *et al.* Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Opt. Lett.* **30**, 468–470 (2005).

185. Popescu, G. *et al.* Imaging red blood cell dynamics by quantitative phase microscopy. *Blood Cells. Mol. Dis.* **41**, 10–16 (2008).

186. Cuche, E., Marquet, P. & Depeursinge, C. Spatial filtering for zero-order and twin-image elimination in digital off-axis holography. *Appl. Opt.* **39**, 4070–4075 (2000).

187. Bianco, V. *et al.* Endowing a plain fluidic chip with micro-optics: a holographic microscope slide. *Light Sci. Appl.* **6**, e17055 (2017).

188. Almoro, P., Pedrini, G. & Osten, W. Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field. *Appl. Opt.* **45**, 8596 (2006).

189. Bao, P., Situ, G., Pedrini, G. & Osten, W. Lensless phase microscopy using phase retrieval with multiple illumination wavelengths. *Appl. Opt.* **51**, 5486–5494 (2012).

190. Min, J. *et al.* Phase retrieval without unwrapping by single-shot dual-wavelength digital holography. *J. Opt.* **16**, 125409 (2014).

191. Ren, Z., Xu, Z. & Lam, E. Y. Autofocusing in digital holography using deep learning. in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXV* vol. 10499 104991V (International Society for Optics and Photonics, 2018).

192. Shimobaba, T., Kakue, T. & Ito, T. Convolutional neural network-based regression for depth prediction in digital holography. *ArXiv180200664 Cs Eess* (2018).

193. Luo, W., Zhang, Y., Göröcs, Z., Feizi, A. & Ozcan, A. Propagation phasor approach for holographic image reconstruction. *Sci. Rep.* **6**, 22738 (2016).

194. Göröcs, Z. *et al.* A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples. *Light Sci. Appl.* **7**, 66 (2018).

195. Wu, Y. *et al.* Deep Learning Enables High-Throughput Analysis of Particle-Aggregation-Based Biosensors Imaged Using Holography. *ACS Photonics* **6**, 294–301 (2019).

196. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* 234–241 (2015). doi:10.1007/978-3-319-24574-4_28.

197. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv E-Prints* **1412**, arXiv:1412.6980 (2014).

198. Goodfellow, I. *et al.* Generative Adversarial Nets. in *Advances in Neural Information Processing Systems 27* 2672–2680 (2014).

199. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-To-Image Translation With Conditional Adversarial Networks. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1125–1134 (2017).

200. Wang, Z. & Bovik, A. C. A universal image quality index. *IEEE Signal Process. Lett.* **9**, 81–84 (2002).

201. Zhang, Y. *et al.* 3D imaging of optically cleared tissue using a simplified CLARITY method and on-chip microscopy. *Sci. Adv.* **3**, e1700553 (2017).

202. Thevenaz, P., Ruttimann, U. E. & Unser, M. A pyramid approach to subpixel registration based on intensity. *IEEE Trans. Image Process.* **7**, 27–41 (1998).

203. Detect SURF features and return SURFPoints object - MATLAB detectSURFFeatures. https://www.mathworks.com/help/vision/ref/detectsurffeatures.html.

204. Rivenson, Y. *et al.* Deep Learning Enhanced Mobile-Phone Microscopy. *ACS Photonics* **5**, 2354–2364 (2018).

205. Moerner, W. E. & Kador, L. Optical detection and spectroscopy of single molecules in a solid. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).

206. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).

207. Betzig, E. *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642–1645 (2006).

208. Hell, S. W. Far-Field Optical Nanoscopy. *Science* **316**, 1153–1158 (2007).

209. Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M. & Vaziri, A. Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat. Methods* **10**, 1013–1020 (2013).

210. Prevedel, R. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* **11**, 727–730 (2014).

211. Tomer, R. *et al.* SPED Light Sheet Microscopy: Fast Mapping of Biological System Structure and Function. *Cell* **163**, 1796–1806 (2015).

212. Ralston, T. S., Marks, D. L., Scott Carney, P. & Boppart, S. A. Interferometric synthetic aperture microscopy. *Nat. Phys.* **3**, 129–134 (2007).

213. Matsushima, K. Formulation of the rotational transformation of wave fields and their application to digital holography. *Appl. Opt.* **47**, D110–D116 (2008).

214. South, F. A. *et al.* Wavefront measurement using computational adaptive optics. *JOSA A* **35**, 466–473 (2018).

215. Nguyen, J. P. *et al.* Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. *Proc. Natl. Acad. Sci.* **113**, E1074–E1081 (2016).

216. Siedentopf, H. & Zsigmondy, R. Uber sichtbarmachung und grö\s senbestimmung ultramikoskopischer teilchen, mit besonderer anwendung auf goldrubingläser. *Ann. Phys.* **315**, 1–39 (1902).

217. Lerner, T. N. *et al.* Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* **162**, 635–647 (2015).

218. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat. Methods* **7**, 339–340 (2010).

219. Abraham, A. V., Ram, S., Chao, J., Ward, E. S. & Ober, R. J. Quantitative study of single molecule location estimation techniques. *Opt. Express* **17**, 23352–23373 (2009).

220. Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027–1036 (2011).

221. Juette, M. F. *et al.* Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples. *Nat. Methods* **5**, 527–529 (2008).

222. Pavani, S. R. P. *et al.* Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Natl. Acad. Sci.* **106**, 2995–2999 (2009).

223. Levoy, M., Ng, R., Adams, A., Footer, M. & Horowitz, M. Light Field Microscopy. in *ACM SIGGRAPH 2006 Papers* 924–934 (ACM, 2006). doi:10.1145/1179352.1141976.

224. Pégard, N. C. *et al.* Compressive light-field microscopy for 3D neural activity recording. *Optica* **3**, 517–524 (2016).

225. Broxton, M. *et al.* Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express* **21**, 25418–25439 (2013).

226. Cohen, N. *et al.* Enhancing the performance of the light field microscope using wavefront coding. *Opt. Express* **22**, 24817–24839 (2014).

227. Abrahamsson, S. *et al.* Fast multicolor 3D imaging using aberration-corrected multifocus microscopy. *Nat. Methods* **10**, 60–63 (2013).

228. Abrahamsson, S. *et al.* MultiFocus Polarization Microscope (MF-PolScope) for 3D polarization imaging of up to 25 focal planes simultaneously. *Opt. Express* **23**, 7734–7754 (2015).

229. Siegel, N., Lupashin, V., Storrie, B. & Brooker, G. High-magnification super-resolution FINCH microscopy using birefringent crystal lens interferometers. *Nat. Photonics* **10**, 802–808 (2016).

230. Shaw, P. J. & Rawlins, D. J. The point-spread function of a confocal microscope: its measurement and use in deconvolution of 3-D data. *J. Microsc.* **163**, 151–165 (1991).

231. Kirshner, H., Aguet, F., Sage, D. & Unser, M. 3-D PSF fitting for fluorescence microscopy: implementation and localization application. *J. Microsc.* **249**, 13–25 (2013).

232. Nguyen, J. P., Linder, A. N., Plummer, G. S., Shaevitz, J. W. & Leifer, A. M. Automatically tracking neurons in a moving and deforming brain. *PLOS Comput. Biol.* **13**, e1005517 (2017).

233. Kato, S. *et al.* Global Brain Dynamics Embed the Motor Command Sequence of Caenorhabditis elegans. *Cell* **163**, 656–669 (2015).

234. Nagy, S., Huang, Y.-C., Alkema, M. J. & Biron, D. *Caenorhabditis elegans* exhibit a coupling between the defecation motor program and directed locomotion. *Sci. Rep.* **5**, 17174 (2015).

235. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy. *Science* **319**, 810–813 (2008).

236. Antipa, N. *et al.* DiffuserCam: lensless single-exposure 3D imaging. *Optica* **5**, 1–9 (2018).

237. Shechtman, Y., Sahl, S. J., Backer, A. S. & Moerner, W. E. Optimal Point Spread Function Design for 3D Imaging. *Phys. Rev. Lett.* **113**, 133902 (2014).

238. Brenner, S. The genetics of Caenorhabditis elegans. *Genetics* **77**, 71–94 (1974).

239. *C. elegans: Methods and Applications*. (Humana Press, 2006).

240. Forster, B., Van De Ville, D., Berent, J., Sage, D. & Unser, M. Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images. *Microsc. Res. Tech.* **65**, 33–42 (2004).

241. Zack, G. W., Rogers, W. E. & Latt, S. A. Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **25**, 741–753 (1977).

242. Mao, X. *et al.* Least squares generative adversarial networks. in *Computer Vision (ICCV), 2017 IEEE International Conference on* 2813–2821 (IEEE, 2017).

243. Shi, J. & Malik, J. Normalized Cuts and Image Segmentation. *IEEE Trans Pattern Anal Mach Intell* **22**, 888–905 (2000).

244. von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416 (2007).

245. Image Stitching. *ImageJ* https://imagej.net/Image_Stitching.

246. US EPA. Particulate matter (PM) basics. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics#PM.

247. Samet, J. *et al.* The carcinogenicity of outdoor air pollution. *Res Rep Health Eff Inst* **140**, 5–114 (2009).

248. Harrison, D., Maggs, R. & Booker, J. UK equivalence programme for monitoring of particulate matter. *Rep. Prod. Bur. Veritas Behalf Defra Devolved Adm. Rep. Ref. No BVAQAD202209DH2396* (2006).

249. Model 804 Handheld Particle Counter | Met One Instruments. http://www.metone.com/particulate-804.php.

250. Fluke 985 Indoor Air Quality Particle Counter. http://en-us.fluke.com/products/hvac-iaq-tools/fluke-985-hvac-iaq.html.

251. TSI Model 3007 hand-held particle counter. http://www.tsi.com/uploadedFiles/_Site_Root/Products/Literature/Spec_Sheets/3007_1930032.pdf.

252. aerosol science & engineering. http://www.aerosols.wustl.edu/Education/default.htm.

253. Orr, C. & Martin, R. A. Thermal precipitator for continuous aerosol sampling. *Rev. Sci. Instrum.* **29**, 129–130 (1958).

254. Broßell, D. *et al.* A thermal precipitator for the deposition of airborne nanoparticles onto living cells—rationale and development. *J. Aerosol Sci.* **63**, 75–86 (2013).

255. Liu, C. *et al.* Transparent air filter for high-efficiency PM2.5 capture. *Nat. Commun.* **6**, 6205 (2015).

256. What is a cleanroom? Cleanroom classifications, class 1, 10, 100, 1,000, 10,000, 100,000, ISO Standard 14644, cleanroom definition. http://www.cleanairtechnology.com/cleanroom-classifications-class.php.

257. Hollander, M., Wolfe, D. A. & Chicken, E. *Nonparametric statistical methods*. (John Wiley & Sons, 2013).

258. Sand Fire Explodes to Over 3,300 Acres in Santa Clarita Area; New Evacuations Ordered in Little Tujunga Canyon. *KTLA* http://ktla.com/2016/07/22/20-acre-fire-burns-near-sand-canyon-in-santa-clarita/ (2016).

259. says, C. H. SCVNews.com | Sand Fire Grows to 3,327 Acres; 200-300 Evacuations | 07-22-2016. http://scvnews.com/2016/07/22/sand-fire-grows-to-2500-acres-evacuations/.

260. Google Maps. https://www.google.com/maps/@33.9465247,-118.3382833,1718m/data=!3m1!1e3.

261. 1,500-Acre La Tuna Canyon Fire Threatens Homes, Prompts 210 Freeway Closure. *KTLA* http://ktla.com/2017/09/01/210-freeway-closed-in-both-directions-due-to-wildfire-threatening-homes-in-tujunga-area/ (2017).

262. Vives, R., Castillo, A. & Tchekmedyian, A. Wildfire brings smoke, ash, unhealthful air quality to parts of Southern California. *Los Angeles Times* (2017).

263. Wu, Y. *et al.* Spatial mapping and analysis of aerosols during a forest fire using computational mobile microscopy. in *Optics and Biophotonics in Low-Resource Settings IV* vol. 10485 104850T (International Society for Optics and Photonics, 2018).

264. Lighty, J. S., Veranth, J. M. & Sarofim, A. F. Combustion Aerosols: Factors Governing Their Size and Composition and Implications to Human Health. *J. Air Waste Manag. Assoc.* **50**, 1565–1618 (2000).

265. Jaworek, A., Czech, T., Sobczyk, A. T. & Krupa, A. Properties of biomass vs. coal fly ashes deposited in electrostatic precipitator. *J. Electrost.* **71**, 165–175 (2013).

266. Chakrabarty, R. K. *et al.* Emissions from the laboratory combustion of wildland fuels: Particle morphology and size. *J. Geophys. Res. Atmospheres* **111**, D07204 (2006).

267. Bari, Md. A. *et al.* Characterisation of particulates and carcinogenic polycyclic aromatic hydrocarbons in wintertime wood-fired heating in residential areas. *Atmos. Environ.* **45**, 7627–7634 (2011).

268. Quality Assurance Air Monitoring Site Information. https://www.arb.ca.gov/qaweb/site.php?s_arb_code=70074.

269. LAX noise management - contour maps. http://www.lawa.org/welcome_lax.aspx?id=1090.

270. FlightStats - Global flight tracker, status tracking and airport information. http://www.flightstats.com/go/Home/home.do.

271. Hudda, N., Gould, T., Hartin, K., Larson, T. V. & Fruin, S. A. Emissions from an international airport increase particle number concentrations 4-fold at 10 km downwind. *Environ. Sci. Technol.* (2014).

272. Windfinder.com - Wind and weather statistic Los Angeles Airport. https://www.windfinder.com/windstatistics/los_angeles_airport.

273. Walton, W. H. & Vincent, J. H. Aerosol instrumentation in occupational hygiene: an historical perspective. *Aerosol Sci. Technol.* **28**, 417–438 (1998).

274. Aerosol Instrumentation » Section 3. http://aerosol.ees.ufl.edu/instrumentation/section03.html.

275. Train support vector machine classifier - MATLAB svmtrain. https://www.mathworks.com/help/stats/svmtrain.html.

276. Grant, M. & Boyd, S. CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx/ (2013).

277. Aimanianda, V. *et al.* Surface hydrophobin prevents immune recognition of airborne fungal spores. *Nature* **460**, 1117-U79 (2009).

278. Srikanth, P., Sudharsanam, S. & Steinberg, R. Bio-aerosols in indoor environment: Composition, health effects and analysis. *Indian J. Med. Microbiol.* **26**, 302 (2008).

279. Codina, R., Fox, R. W., Lockey, R. F., DeMarco, P. & Bagg, A. Typical levels of airborne fungal spores in houses without obvious moisture problems during a rainy season in Florida, USA. *J. Investig. Allergol. Clin. Immunol.* **18**, 156–162 (2008).

280. de Ana, S. G., Torres-Rodríguez, J. M., Ramírez, E. A., García, S. M. & Belmonte-Soler, J. Seasonal distribution of Alternaria, Aspergillus, Cladosporium and Penicillium species isolated in homes of fungal allergic patients. *J. Investig. Allergol. Clin. Immunol.* **16**, 357–363 (2006).

281. Lee, T. *et al.* Culturability and concentration of indoor and outdoor airborne fungi in six single-family homes. *Atmospheric Environ. Oxf. Engl. 1994* **40**, 2902–2910 (2006).

282. Douglas, P., Robertson, S., Gay, R., Hansell, A. L. & Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int. J. Hyg. Environ. Health* **221**, 134–173 (2018).

283. Kim, K.-H., Kabir, E. & Jahan, S. A. Airborne bioaerosols and their impact on human health. *J. Environ. Sci.* **67**, 23–35 (2018).

284. Fischer, G., Albrecht, A., Jäckel, U. & Kämpfer, P. Analysis of airborne microorganisms, MVOC and odour in the surrounding of composting facilities and implications for future investigations. *Int. J. Hyg. Environ. Health* **211**, 132–142 (2008).

285. Hryhorczuk, D. *et al.* Bioaerosol emissions from a suburban yard waste composting facility. *Ann. Agric. Environ. Med. AAEM* **8**, 177–185 (2001).

286. Pearson, C. *et al.* Exposures and Health Outcomes in Relation to Bioaerosol Emissions From Composting Facilities: A Systematic Review of Occupational and Community Studies. *J. Toxicol. Environ. Health B Crit. Rev.* **18**, 43–69 (2015).

287. D'Amato, G. *et al.* Allergenic pollen and pollen allergy in Europe. *Allergy* **62**, 976–990.

288. Dedesko, S., Stephens, B., Gilbert, J. A. & Siegel, J. A. Methods to assess human occupancy and occupant activity in hospital patient rooms. *Build. Environ.* **90**, 136–145 (2015).

289. WHO, Ambient (outdoor) air quality and health. http://www.who.int/mediacentre/factsheets/fs313/en/.

290. Walt, D. R. & Franz, D. R. Peer Reviewed: Biological Warfare Detection. *Anal. Chem.* **72**, 738–A (2000).

291. Mandal, J. & Brandl, H. Bioaerosols in Indoor Environment - A Review with Special Reference to Residential and Occupational Locations. *Open Environ. Biol. Monit. J.* **4**, 83–96 (2011).

292. Usachev, E. V., Tam, A. M., Usacheva, O. V. & Agranovski, I. E. The sensitivity of surface plasmon resonance based viral aerosol detection. *J. Aerosol Sci.* **76**, 39–47 (2014).

293. Halsby, K. D. *et al.* The Epidemiology of Q Fever in England and Wales 2000–2015. *Vet. Sci.* **4**, E28 (2017).

294. Haig, C. W., Mackay, W. G., Walker, J. T. & Williams, C. Bioaerosol sampling: sampling mechanisms, bioefficiency and field studies. *J. Hosp. Infect.* **93**, 242–255 (2016).

295. Wu, Y., Shen, F. & Yao, M. Use of gelatin filter and BioSampler in detecting airborne H5N1 nucleotides, bacteria and allergens. *J. Aerosol Sci.* **41**, 869–879 (2010).

296. Rodriguez-Riano, T. & Dafni, A. A new procedure to asses pollen viability. *Sex. Plant Reprod.* **12**, 241–244 (2000).

297. Heslop-Harrison, J. & Heslop-Harrison, Y. Evaluation of pollen viability by enzymatically induced fluorescence; intracellular hydrolysis of fluorescein diacetate. *Stain Technol.* **45**, 115–120 (1970).

298. National Allergy Forecast & Info About Allergies | Pollen.com. https://www.pollen.com/.

299. Kaye, P. H. *et al.* Single particle multichannel bio-aerosol fluorescence sensor. *Opt. Express* **13**, 3583–3593 (2005).

300. Reyes, F. L. *et al.* Bio-aerosol fluorescence sensor. *Field Anal. Chem. Technol.* **3**, 240–248 (1999).

301. Eversole, J. D., Hardgrove, J. J., Cary, W. K., Choulas, D. P. & Seaver, M. Continuous, rapid biological aerosol detection with the use of UV fluorescence: Outdoor test results. *Field Anal. Chem. Technol.* **3**, 249–259 (1999).

302. Ruske, S. *et al.* Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer. *Atmos Meas Tech* **10**, 695–708 (2017).

303. Leśkiewicz, M. *et al.* Improved real-time bio-aerosol classification using Artificial Neural Networks. *Atmospheric Meas. Tech. Discuss.* 1–19 (2018) doi:https://doi.org/10.5194/amt-2018-86.

304. Johnson, B. N. & Mutharasan, R. Biosensing using dynamic-mode cantilever sensors: A review. *Biosens. Bioelectron.* **32**, 1–18 (2012).

305. Kovář, D., Farka, Z. & Skládal, P. Detection of Aerosolized Biological Agents Using the Piezoelectric Immunosensor. *Anal. Chem.* **86**, 8680–8686 (2014).

306. Usachev, E. V., Usacheva, O. V. & Agranovski, I. E. Surface plasmon resonance-based real-time bioaerosol detection. *J. Appl. Microbiol.* **115**, 766–773 (2013).

307. Schwarzmeier, K., Knauer, M., Ivleva, N. P., Niessner, R. & Haisch, C. Bioaerosol analysis based on a label-free microarray readout method using surface-enhanced Raman scattering. *Anal. Bioanal. Chem.* **405**, 5387–5392 (2013).

308. Grow, A. E., Wood, L. L., Claycomb, J. L. & Thompson, P. A. New biochip technology for label-free detection of pathogens and their toxins. *J. Microbiol. Methods* **53**, 221–233 (2003).

309. DeAngelis, K. M. *et al.* PCR Amplification-Independent Methods for Detection of Microbial Communities by the High-Density Microarray PhyloChip ▿. *Appl. Environ. Microbiol.* **77**, 6313–6322 (2011).

310. Verma, J., Saxena, S. & Babu, S. G. ELISA-Based Identification and Detection of Microbes. in *Analyzing Microbes* 169–186 (Springer, Berlin, Heidelberg, 2013). doi:10.1007/978-3-642-34410-7_13.

311. Cha, S. *et al.* Metagenomic Analysis of Airborne Bacterial Community and Diversity in Seoul, Korea, during December 2014, Asian Dust Event. *PLOS ONE* **12**, e0170693 (2017).

312. Thomas, T., Gilbert, J. & Meyer, F. Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* **2**, 3 (2012).

313. Mandracchia, B. *et al.* Holographic microscope slide in a spatio-temporal imaging modality for reliable 3D cell counting. *Lab. Chip* **17**, 2831–2838 (2017).

314. Solomon, W. R., Burge, H. A. & Muilenberg, M. L. Allergen carriage by atmospheric aerosol. I. Ragweed pollen determinants in smaller micronic fractions. *J. Allergy Clin. Immunol.* **72**, 443–447 (1983).

315. Habenicht, H. A., Burge, H. A., Muilenberg, M. L. & Solomon, W. R. Allergen carriage by atmospheric aerosol: II. Ragweed-pollen determinants in submicronic atmospheric fractions. *J. Allergy Clin. Immunol.* **74**, 64–67 (1984).

316. Fernandez-Caldas, E. *et al.* Immunochemical Demonstration of Red Oak Pollen Aeroallergens outside the Oak Pollination Season. *Grana* **28**, 205–209 (1989).

317. Miccio, L. *et al.* Dynamic DIC by digital holography microscopy for enhancing phase-contrast visualization. *Biomed. Opt. Express* **2**, 331–344 (2011).

318. Bishop, C. *Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn*. (Springer, New York, 2007).

319. CampusPlantsSearchResults – Mildred E. Mathias Botanical Garden. https://www.botgard.ucla.edu/campusplantssearchresults/.

320. Olivo-Marin, J.-C. Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35**, 1989–1996 (2002).

321. Xu, L., Oja, E. & Kultanen, P. A new curve detection method: Randomized Hough transform (RHT). *Pattern Recognit. Lett.* **11**, 331–338 (1990).