Title
Deep Learning-enabled Cross-modality Image Transformation and Early Bacterial Colony Detection

Author
Wang, Hongda

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Deep Learning-enabled Cross-modality Image Transformation and Early Bacterial Colony Detection

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical and Computer Engineering

by

Hongda Wang

2020

ABSTRACT OF THE DISSERTATION

Deep Learning-enabled Cross-modality Image Transformation

and Early Bacterial Colony Detection

by

Hongda Wang

Doctor of Philosophy in Electrical & Computer Engineering

University of California, Los Angeles, 2020

Professor Aydogan Ozcan, Chair

Recent developments of deep learning-enabled image transformation and object detection in microscopic images has revolutionized traditional computational imaging techniques and outperformed many digital image processing algorithms in both speed and quality. This dissertation introduces a set of novel deep learning techniques for cross-modality image super-resolution, virtual histological staining, and early bacterial colony using time-lapsed coherent microscopic images. This dissertation first introduces a deep learning-based method to correct distortions introduced by mobile-phone-based microscopes is introduced, which facilitates the production of high-resolution, denoised and color-corrected images, matching the performance of benchtop microscopes with high-end objective lenses, also extending their limited depth-of-field.

Inspired mobile-phone microscope to benchtop microscope image transformation, a deep learning-enabled super-resolution framework across different fluorescence microscopy modalities

is also demonstrated. Using this framework, the resolution of wide-field images acquired with low-numerical-aperture (NA) objectives were improved to match the resolution that is acquired using high-NA objectives. The framework was further applied to cross-modality super-resolution transformation of confocal microscopy images to match the resolution acquired with a stimulated emission depletion (STED) microscope, and transformation of total internal reflection fluorescence (TIRF) microscopy images of subcellular structures within cells and tissues to match the results obtained with a TIRF-based structured illumination microscope.

The similar cross-modality image transformation framework can also transform autofluorescence images of unlabeled tissue sections into the equivalence of the bright-field images captured with histologically stained versions of the same samples. A blind comparison, by board-certified pathologists, of this virtual staining method and standard histological staining using microscopic images of human tissue sections of the salivary gland, thyroid, kidney, liver, and lung, and involving different types of stain, showed no major discordances.

Other than image transformation, a deep learning-based live bacteria detection system was also developed which periodically captures coherent microscopy images of bacterial growth inside a 60-mm-diameter agar plate and analyses these time-lapsed holograms for the rapid detection of bacterial growth and the classification of the corresponding species. This system shortens the detection time of *Escherichia coli* and total coliform bacteria in water samples by >12 h compared to the Environmental Protection Agency (EPA)-approved methods, achieved a limit of detection (LOD) of ~1 colony forming unit (CFU)/L in $\leqslant$9 h of total test time. This platform is highly suitable for integration with the existing methods currently used for bacteria detection on agar plates.

The dissertation of Hongda Wang is approved.

Pei-Yu Chiou

Sam Emaminejad

Chee Wei Wong

Aydogan Ozcan, Committee Chair

University of California, Los Angeles

2020

Table of Contents

Acknowledgements

I would first like to thank my Ph.D. advisor, Prof. Aydogan Ozcan, whose expertise and mentoring was invaluable throughout my Ph.D. study and research. I feel greatly honored to be able to join Prof. Ozcan's lab where I gained research knowledge and skills during the years working here. More importantly, Prof. Ozcan has helped me build a high standard of scholarship and a wide scope for evaluating the significance of research outcomes. Prof. Ozcan has also been a personal role model for me. His dedication in pursuit of excellence as well as his innovative insights to science and engineering keep me inspired for every moment.

I would also like to thank my dissertation committee members, Prof. Pei-Yu Chiou, Prof. Sam Emaminejad, and Prof. Chee Wei Wong, for their precious advice in improving my research skills and the preparation of this dissertation.

During the years of my Ph.D. study, the great colleagues and collaborators are of exceptional importance, without whom many of my research projects would not be made possible. I would like to thank Dr. Yair Rivenson, Dr. Laurent Bentolila, Dr. Yibo Zhang, Dr. Yichen Wu, Dr. Zoltán Göröcs, Dr. Hatice Ceylan Koydemir, Dr. Wei Luo, Dr. Michael Lake, Dr. Zach Ballard, Dr. Daniel Shir, Dr. Hyouarm Joung, Dr. Aniruddha Ray, Dr. Comert Kural, Dr. Jonathan E. Zuckerman, Dr. W. Dean Wallace, Dr. Thomas Chong, Dr. Anthony E. Sisk, Dr. Lindsey M. Westbrook, Bijie Bai, Miu Tamamitsu, Tairan Liu, Yi Luo, Zhensong Wei, Kevin de Haan, Calvin Brown, Derek Tseng, Deniz Mengu, Yunzhe Qiu, Yiyin Jin, Harun Gunaydin, Kyle Liang, Zhengshuang Ren, and Ronald Gao, among many others for the support during my Ph.D. research.

Finally, I would like to express my sincere gratitude to my family for their unconditional support. Their love, understand, and encouragement are the irreplaceable sources of power that accompanied me through the highs and lows during this adventure of Ph.D. study and research.

Vita

Hongda Wang received his B.S. degree from School of Physics, Peking University in Beijing, China in 2015. After that, he joined Prof. Aydogan Ozcan's Bio- and Nano-Photonics Lab at Electrical & Computer Engineering Department, UCLA to pursue a Ph.D. degree. Hongda's research focuses on computational microscopy and deep learning for biomedical applications.

During his education at UCLA, Hongda received the Keystone Symposia: Future of Science Fund Scholarship, UCLA Photonics Scholar Fellowship, UCLA Graduate Division Fellowship, and Dissertation Year Fellowship. Hongda has co-authored 16 high-impact journal publications and published/presented in over 35 conference meetings.

Selected publications

1. **H. Wang**, Z. Göröcs, W. Luo, Y. Zhang, Y. Rivenson, L. A. Bentolila, and A. Ozcan, "Computational out-of-focus imaging increases the space–bandwidth product in lens-based coherent microscopy," Optica **3**, 1422–1429 (2016).
2. **H. Wang**, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nat. Methods **16**, 103–110 (2019).
3. **H. Wang**, H. Ceylan Koydemir, Y. Qiu, B. Bai, Y. Zhang, Y. Jin, S. Tok, E. C. Yilmaz, E. Gumustekin, Y. Rivenson, and A. Ozcan, "Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning," Light Sci. Appl. **9**, 118 (2020).
4. Y. Rivenson, **H. Wang**, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nat. Biomed. Eng. **3**, 466 (2019).
5 Y. Rivenson, H. Ceylan Koydemir, **H. Wang**, Z. Wei, Z. Ren, H. Günaydın, Y. Zhang, Z. Göröcs, K. Liang, D. Tseng, and A. Ozcan, "Deep Learning Enhanced Mobile-Phone Microscopy," ACS Photonics (2018).

Chapter 1 Introduction to deep learning microscopy

1.1 Optical microscopy as a fundamental tool for research and healthcare

Optical microscopy is one of the fundamental tools for studying physiological process in life science research and has enabled numerous key discoveries. It is also regarded as the gold standard method for the medical diagnosis of various diseases. However, the performance of an optical microscope such as spatial and temporal resolution is limited by the physical hardware and the traditional direct observing method, which cannot meet the ever-growing demands among researchers and healthcare professionals. To overcome these difficulties, the concept of computational microscopy has been developed over the past few decades, which uses computational post-processing algorithms to exploit the imaging capabilities of a microscope and visualize hard to observe information such as phase [1,2] and birefringence [3,4]. More recently, the rapid development of deep learning techniques brings new opportunities in the computational imaging field and enables new applications such as image super-resolution [5–8], virtual histological staining [9–11], holographic image reconstruction [12–14], etc. Besides improved processing speed and image qualities over the traditional computational algorithms, most importantly, the deep learning techniques do not need a physical model of the image formation process, therefore can achieve cross-modality image transformation where physical models are extremely hard to build, if not impossible. This non-physics-based image transformation capability opens up a whole new range of applications, such as cross-modality image transformation from confocal microscopy to stimulated emission depletion (STED) microscopy, virtual histological staining of unlabeled tissue sections, and many more. The deep learning framework is also well-suited for processing high-dimension data, which will be demonstrated by early detection and

classification of bacterial colonies using image stacks that contains spatial-time-amplitude/phase information.

In this dissertation, I will start with the pixel super-resolution algorithm for coherent microscopic image reconstruction as an example of the traditional computational techniques. Then I will focus on deep learning techniques developed in recent years for super-resolution imaging and cross-modality transformations, as well as early detection of live bacteria using time-lapsed coherent imaging.

In Chapter 1, I will introduce the background of computational imaging and demonstrated out-of-focus pixel super-resolution (OFI-PSR) based image super-resolution techniques developed at the beginning of my PhD training. Then I will introduction deep learning microscopy concepts and recent developments in this field and demonstrated deep learning enhanced mobile-phone microscopy. Part of this chapter has been published in

- H. Wang, Z. Göröcs, W. Luo, Y. Zhang, Y. Rivenson, L. A. Bentolila, and A. Ozcan, "Computational out-of-focus imaging increases the space–bandwidth product in lens-based coherent microscopy," Optica 3, 1422–1429 (2016).
- Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Günaydın, Y. Zhang, Z. Göröcs, K. Liang, D. Tseng, and A. Ozcan, "Deep Learning Enhanced Mobile-Phone Microscopy," ACS Photonics (2018).

1.2 Computational out-of-focus imaging increases the space-bandwidth product in lens-based coherent microscopy

Introduction

Although modern microscope objective-lenses can achieve high-resolution imaging with relatively large fields-of-view (FOV), they are inherently designed to provide a match to human eye rather than to charge-coupled device (CCD) or complementary metal-oxide-semiconductor (CMOS) based cameras, which appeared in recent decades as common microscope accessories. The space-bandwidth product (SBP) of an optical system is defined by the FOV of the imaging platform divided by the area of a resolvable spot, which is determined by the spatial resolution of the imager [15] and in this sense it is fundamentally tied to the signal-to-noise ratio of the optical imaging system. In case the spatial resolution exhibits significant variations across the claimed FOV of the imaging system, e.g., due to aberrations etc., SBP can be estimated by defining sub-regions of the FOV, each with a uniform resolution. For a coherent imaging system, both phase and amplitude channels would independently contribute to the SBP, and for example a 10× objective-lens with a numerical aperture (NA) of 0.3 and a field number (FN) of 26.5 mm can achieve, if corrected for aberrations, a total SBP of approximately 14 million at an illumination wavelength of 532 nm. However, due to the signal readout mechanism and imaging speed requirements, most cameras that are used in optical microscopes are designed with limited number of pixels, e.g., 1-4 megapixels, which sets a practical limitation for the overall SBP of the microscopic imaging system (see **Figure 1.1**). This gap between objective-lenses and opto-electronic sensor chips is in general bridged by matching the optical resolution to the effective pixel size of the imaging configuration, which results in a major sacrifice of the FOV. For example, the use of the same 10×/0.3NA objective-lens with a commonly used 1.45-megapixel CCD imager

(QIClick Monochrome, QImaging, Surrey, BC, Canada) would necessitate at least a 1× camera adaptor to effectively reduce the pixel size by 10-fold and match the resolution of the objective-lens to the CCD chip. This strategy would unfortunately waste >89% of the objective-lens FOV and therefore result in sub-optimal use of the SBP of the microscopic imaging system.

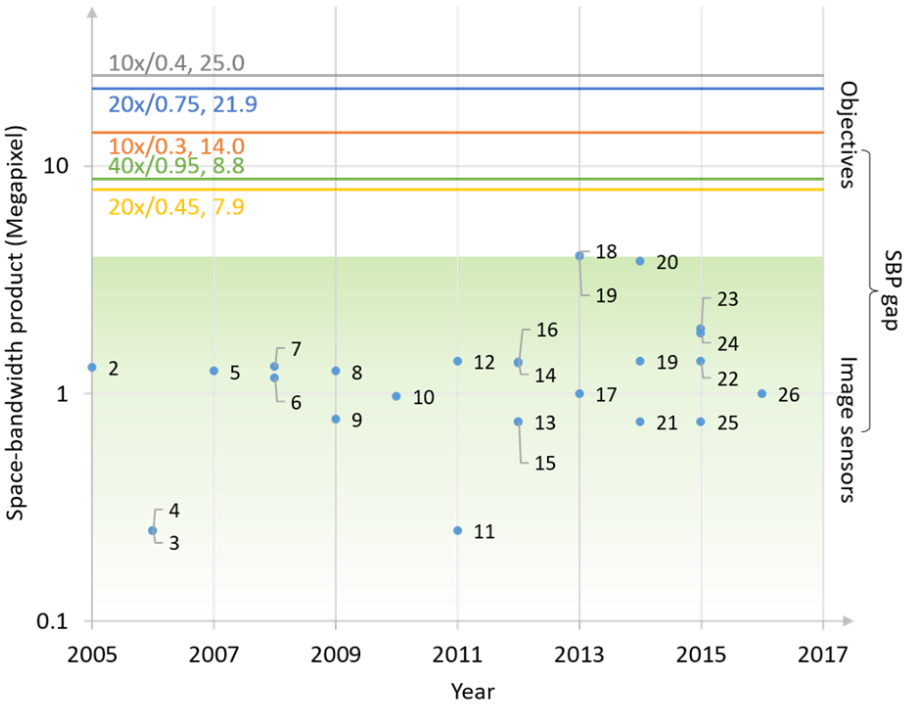

Figure 1.1 Space-bandwidth product (SBP) gap between microscope objectives and image sensors that are employed in coherent imaging experiments. Coherent microscopy and digital holographic imaging fields have been using CCD/CMOS image sensors typically with less than 4 million pixels during the past decade (green highlighted area), while the existing objective-lenses can achieve a SBP of ~8-25 million (solid lines). I will term this practical mismatch between the pixel-counts of camera sensor chips and the SBPs of conventional objective-lenses as the SBP gap in coherent microscopic imaging. Each blue point refers to the pixel-count of the image sensor reported in a publication indicated by the reference number next to it. The SBPs reported for these objective-lenses include both the phase and amplitude channels and assume a coherent illumination at 532 nm wavelength.

In fact, a survey of coherent imaging and digital holographic microscopy related publications from the past decade clearly illustrates this mismatch between the pixel-counts of the utilized image sensor chips and the SBPs of conventional microscope objectives, as summarized in Figure 1.1. [16–40] We refer to this practical mismatch as the "SBP gap" in coherent microscopy systems. To address this gap, here I introduce a new wide-field and high-resolution computational imaging method that best utilizes the SBP of a microscope objective by bridging the gap between digital cameras and objective-lenses. For this goal, unlike traditional microscope designs, I first add a demagnification camera adaptor (e.g., 0.35×) to match the CCD/CMOS image sensor area to the FOV of the objective-lens. This demagnification operation, although increases the sample FOV, reduces the image resolution due to inadequate sampling and results in spatial aliasing and pixelation. To mitigate this limitation, I employ a pixel super-resolution algorithm that uses a few out-of-focus images of the sample to recover a high-resolution complex image of the specimen and significantly increase the overall SBP of the microscope. Conventional pixel super-resolution (PSR) methods restore high-frequency signals from a stack of undersampled images, each with a sub-pixel lateral displacement. Such PSR methods are implemented by either laterally shifting the sample [41] or shifting the sensor chip inside a camera. The former method needs high precision motorized stages and may have anisotropic resolution due to uneven sub-pixel movements/shifts. The latter, on the other hand, requires a specialized camera (e.g., DP80, Olympus [42]) with a built-in pixel-shifting mechanism and a Peltier cooling device. Both of these PSR implementations inevitably complicate the mechanical design of the microscope and increase the hardware costs.

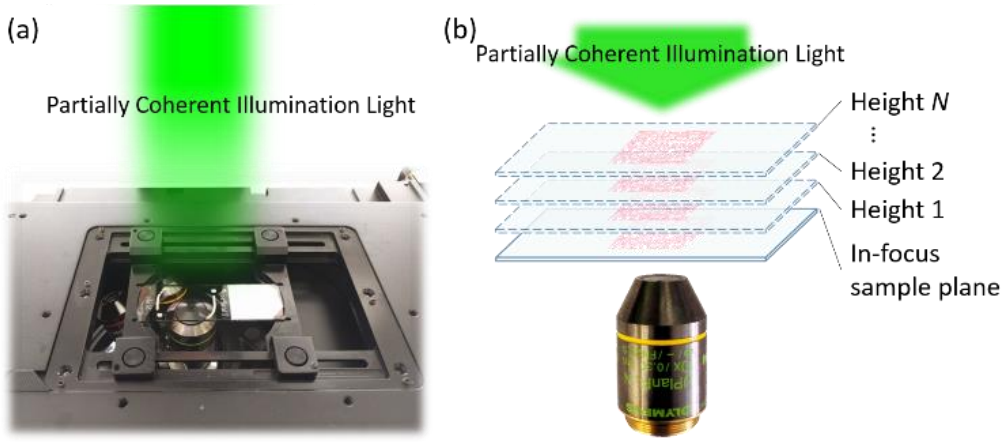

Figure 1.2 The out-of-focus coherent microscopy setup. (a) Quasi-monochromatic illumination at 532 nm wavelength with ~2 nm bandwidth is used for illumination. A 0.35× demagnification camera adaptor was introduced to increase the FOV by ~8 fold, getting close to the FOV limit of the objective-lens. (b) A stack of out-of-focus images is captured by vertically moving the objective-lens, which is then used to digitally recover a wide-field and high-resolution complex image of the sample, including both phase and amplitude channels. Typically, N~3-5.

Using a stack of out-of-focus images of the sample, I developed a pixel super-resolution framework to create high-resolution and wide-field microscopic images of specimen, both amplitude and phase, with minimal changes to a conventional bright-field microscope, providing much better utilization of the large SBP of a microscope objective-lens. The feasibility of this approach, which is termed as out-of-focus imaging-based pixel super-resolution (OFI-PSR), is demonstrated by reconstructing a resolution test-target as well as various biological samples, including e.g., blood samples and Papanicolaou smears. The same imaging technique can also be extended to 3D objects assuming that shadowing artifacts due to object thickness and optical density do not create major limitations. To achieve the same SBP that is inherently limited by the objective-lens, my approach requires ~5-6 and ~3 fold less number of images when compared to traditional off-axis and phase-shifting digital holographic microscopy techniques, respectively (**Table 1.1**). This unique technique would be useful to optimize the throughput and SBP of lens-

based coherent imaging platforms and might inspire new microscopy systems that benefit from the built-in auto-focusing process of an automated scanning microscope to further increase its SBP.

Materials and methods

Experimental setup

The OFI-PSR method is demonstrated experimentally using a conventional bright-field microscope (IX73, Olympus Corporation, Tokyo, Japan). **Figure 1.2** depicts the objective-lens based out-of-focus coherent imaging setup. A fiber-coupled wavelength-tunable light source (WhiteLase-Micro, model VIS, Fianium Ltd, Southampton, UK) is used to provide the illumination. This tunable light source is set to 532 nm with ~2 nm bandwidth. The partially coherent characteristic of the light source allows us to treat each out-of-focus image as an in-line transmission hologram of the sample, while also avoiding any interference from objects outside of the sample plane. A CCD-based image sensor (QIClick Monochrome, QImaging, Surrey, BC, Canada) with a pixel-count of 1.45 million and a pixel size of 6.45 μm is used to capture the out-of-focus transmission images. In this microscopic imaging system, I also introduced a demagnification factor of 0.35× by adding a camera adapter (Olympus Part #U-TV0.35xC-2) to increase the FOV by ~8 fold, getting close to the FOV limit of the objective-lens. With this demagnification, the sample FOV becomes 4.6 mm^2 using a 10×/0.3NA objective-lens (FN = 26.5 mm) and 1.1 mm2 using a 20×/0.45NA objective-lens (FN = 22 mm). Note that either the sample or the objective-lens can be scanned vertically to capture the required out-of-focus images. In my experimental implementation, the objective-lens was scanned vertically, and to investigate the optimum number of out-of-focus images, I used an exploratory depth imaging range of ~40 μm to ~400 μm with respect to the sample plane, with an axial step size of ~15 μm (see **Figure 1.2**b).

As will be demonstrated in the Results and Discussion Section, ~3-5 out-of-focus measurements separated by 30 μm are sufficient to reconstruct high quality images.

Sample preparation

I validated OFI-PSR by imaging a standard 1951 USAF resolution test-target as well as unstained Papanicolaou (Pap) smears and blood samples. Pap smears are prepared using ThinPrep® method (Hologic, Massachusetts, USA). The human blood smear is acquired from Carolina (item no. 31-7374). Since I used existing and anonymous specimen, where no subject related information is linked or can be retrieved, these experiments were exempt from human subject research related regulations.

OFI-PSR algorithm

First, I assume that the quasi-monochromatic light field right after the object plane can be expressed as $o(x,y)=1+s(x,y)$ where $s(x,y)$ is the object transmission field. The spatial Fourier transform of the sampled intensity $I_{\text{sampled},k}(f_x, f_y)$ for each out-of-focus measurement (k) can be written as:

$$I_{\text{sampled},k} = \sum_{u,v=0,\pm1,\pm2,\cdots} [\delta_{uv} + H_k^*(0,0)\cdot H_{uv,k}\cdot S_{uv} + H_k(0,0)\cdot (H_{uv,k}^{-}\cdot S_{uv}^{-})^* + SS_{uv,k}]\cdot P_{uv},$$

where the superscript '-' represents $(-f_x,-f_y)$ instead of (f_x, f_y), the asterisk stands for the complex conjugate operation and $H_k(f_x, f_y)$ is the free space transfer function, implemented between the kth out-of-focus sample plane and the in-focus sample plane, separated by z_k, i.e.:

$$H_k(f_x, f_y) = \begin{cases} \exp\left[j \cdot 2\pi \frac{n \cdot z_k}{\lambda} \sqrt{1 - \left(\frac{\lambda}{n} f_x\right)^2 - \left(\frac{\lambda}{n} f_y\right)^2} \right] & \left(f_x^2 + f_y^2 \le \left(\frac{\mathrm{NA}}{\lambda}\right)^2 \right). \\ 0 & \text{Otherwise} \end{cases} \quad (1.2)$$

[1,43,44]. Note that whether it is the sample or the objective-lens that is moved vertically to create out-of-focus images, in my notation z_k refers to the relative depth shift between the in-focus sample plane and a given out-of-focus plane, *k* (**Figure 1.2**b). Each term with the subscript "uv" in Equation (1.1) represents spatial aliasing related replicas, i.e.:

$$F_{uv} = F(f_x - \frac{u}{\Delta x}, f_y - \frac{v}{\Delta y}), \quad (1.3)$$

where Δx and Δy refer to the effective pixel pitch/period (along x and y, respectively) at the focal plane of the objective-lens. In Equation(1.2), P_{uv} refers to the 2D Fourier transform of the "effective pixel function" of the image sensor chip that is projected onto the sample plane that is in focus, and this 2D effective pixel function represents the intensity responsivity distribution of a single pixel at the in-focus sample plane of the microscope. S_{uv} refers to the spatial frequency spectrum of the object scattering field, and S_{00} is the target of the reconstruction algorithm. One should note that both S_{uv} and P_{uv} in Equation(1.1) are independent of the separation between different out-of-focus planes since the illumination wavelength remains unchanged. The last term in Equation(1.1), SS_k, represents the self-interference resulting from out-of-focus imaging related diffraction, which can be written as $SS_k = \Gamma_{f_x, f_y}\{H_k \cdot S_k\}$, where Γ_{f_x, f_y} refers to the 2D autocorrelation operation. [43,45]

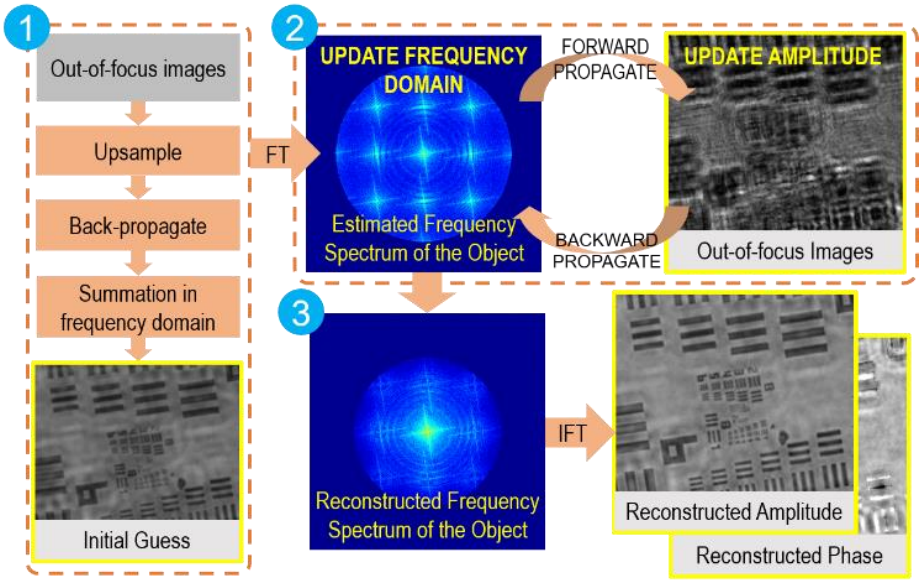

Figure 1.3 Schematic diagram of the OFI-PSR algorithm.

To recover a high-resolution image of the complex object field based on out-of-focus imaging, as depicted in **Figure 1.3** and detailed below, the OFI-PSR algorithm consists of two stages: (I) generation of an initial object guess, and (II) iterative refinement and reconstruction of the complex object in the frequency spectrum.

Stage I: Generation of the initial guess

An initial guess of the frequency spectrum of the object is generated through a three-step procedure. First, each out-of-focus intensity image is upsampled by n fold (e.g., 4-6). This procedure does not introduce any new information but extends the frequency domain window. In the second step, each upsampled out-of-focus image is digitally back-propagated to the in-focus sample plane of the objective lens. In terminology of this thesis, the wave propagation from the in-focus sample plane to an out-of-focus sample plane is denoted as the forward-propagation, and the inverse process from an out-of-focus sample plane to the in-focus sample plane is denoted as backward-propagation. In the final step of this Stage I of OFI-PSR, I sum up all the back-

propagated complex fields calculated from different out-of-focus images and generate an initial guess of the object's spatial frequency spectrum.

Stage II: Iterative frequency spectrum refinement

In the second stage, I use an iterative algorithm to refine the object reconstruction and eliminate aliasing related spatial artifacts. As depicted in **Figure 1.3**, the current iteration (*i*) of the object estimation (o^i) is first forward-propagated to each out-of-focus object plane using the angular spectrum method [15], yielding an estimated out-of-focus image $h^i_{\text{forward},k}$ for the *i*-th iteration, where *k* represents the *k*-th out-of-focus measurement. At the next step of the algorithm, the low resolution raw measurement at the kth out-of-focus measurement plane is convolved with the 2D effective pixel function of the sensor-array, which I assumed to be a Gaussian with a FWHM that is a quarter of the pixel pitch, and the result is used to update the amplitude of $h^i_{\text{forward},k}$, with a relaxation factor of e.g., 0.5, while keeping the phase unchanged. [43] This updated field, $h^{i+1}_{\text{forward},k}$, is then back-propagated to the in-focus sample plane, yielding $o^{i+1}_{\text{backward},k}$, which is used to update the object estimation o^i in the spatial frequency domain, also using a relaxation factor (e.g., ~0.5). Before this update, $o^{i+1}_{\text{backward},k}$ is also filtered by a spatial frequency mask defined by the passband of the coherent imaging system based on the NA of the objective-lens to avoid amplification of high frequency noise during each iteration cycle. After each out-of-focus measurement $I_{\text{sampled},k}$ has been utilized in a given iteration, the object field estimation is updated from o^i to o^{i+1}, and typically I use $i \sim 100$ iterations as part of Stage II.

Estimation of relative axial positions of out-of-focus images

The axial position of each out-of-focus image can be determined by digital auto-focusing algorithms. [41,46] However, such algorithms tend to perform poorly in case of severely undersampled images and are sensitive to noise caused by the interference fringes arising from unwanted objects (e.g., dust, etc.) residing on the optical elements within the beam path. To address this problem, I used an iterative refinement process after obtaining the initial out-of-focus heights through standard auto-focusing algorithms [41,47,48]. Using these initial height estimates, the object field is first digitally propagated to each out-of-focus sample plane, denoted as $o(x, y, z_k^{(i)})$. At height $z_k^{(i)}$, I propagate $o(x, y, z_k^{(i)})$ around the estimation point, searching for a position $z_k^{(i+1)}$ where the correlation of $|o(x, y, z_k^{(i+1)})|^2$ and $I_{\text{sampled},k}(x, y)$ is the largest. Then $z_k^{(i)}$ is replaced with $z_k^{(i+1)}$, updating all the height values corresponding to the out-of-focus measurements. This algorithm converges rapidly and requires about 5 iterations, where the termination criterion is set to be:

$$\frac{1}{N}\sum_{k=1}^{N}| z_k^{(i+1)} - z_k^{(i)} |< \varepsilon, \tag{1.4}$$

where ε is the error tolerance, usually ~0.05 μm.

Through my experiments, I found out that axial step sizes of ~15-60 μm between successive out-of-focus images do not show noticeable differences in the OFI-PSR reconstruction results.

Computation platform for the implementation of OFI-PSR algorithm

The OFI-PSR algorithm is implemented in MATLAB (Version R2016a, MathWorks, Natick, MA, USA) on a desktop computer, equipped with a 3.60 GHz central processing unit (Intel Core i7-4970) and 16 GB of random-access memory. For a stack of N = 5 out-of-focus images with

512×512 pixels each, i.e., covering about 0.8 mm2 field-of-view, one iteration takes approximately 5.8 seconds with an up-sampling factor of 6, such that the total OFI-PSR reconstruction routine finishes within 10 minutes. In my proof-of-concept implementation, the OFI-PSR algorithm was executed sequentially on a CPU, and one can expect a significant reduction in the computation time (e.g., 10-20 fold) with the help of GPUs (graphics processing units) and parallel computing [49].

Results and discussion

The physical basis of the technique relies on the relative changes of out-of-focus images with respective to the image sampling grid as a function of the sample-to-focus distance. There are two main factors affecting the resolution of the reconstructed images using OFI-PSR: signal-to-noise ratio (SNR) of each out-of-focus image and the spatial sampling rate. Poor SNR limits the resolution by affecting the detection of high-order and lower energy interference patterns in each out-of-focus image and reduces the contrast. As for the effective pixel pitch for spatial sampling at the focal plane of the objective-lens, after taking into account the overall magnification of the coherent optical system, it has a relatively large sampling period, which causes severe undersampling in each out-of-focus measurement, in return for a significantly increased sample FOV. As will be detailed next, OFI-PSR not only recovers the phase information of the sample by using a set of intensity-only out-of-focus images, but also performs anti-aliasing by utilizing the strong sensitivity of the coherent transfer function (H_k) to the sample-to-focus distance and reconstructs a pixel super-resolved image of the complex object field, significantly increasing the SBP of a coherent microscopy system.

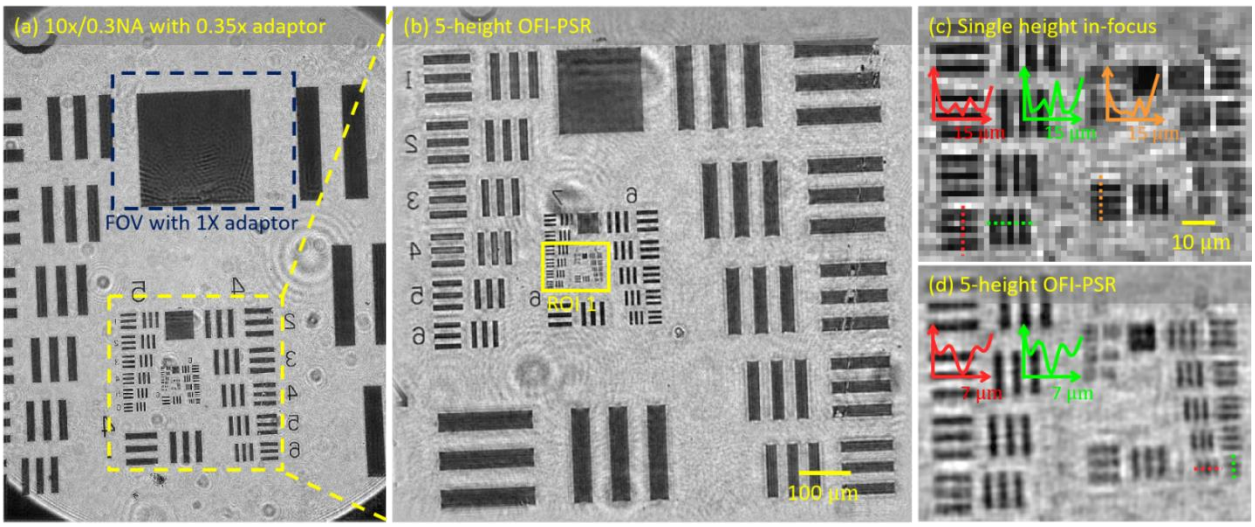

Figure 1.4 OFI-PSR reconstruction results of a resolution test-target using N = 5 out-of-focus images. Microscope objective-lens: 10×/0.3NA; camera adapter: 0.35×; illumination wavelength: 532 nm. (a) Full FOV OFI-PSR reconstruction of ~4.6 mm2 FOV. (b) Zoom-in of (a). (c) Single-height in-focus image of ROI 1 indicates severe spatial undersampling with a lateral half-pitch resolution of ~2.2 μm. (d) OFI-PSR reconstruction result for the same ROI 1 shows a significantly improved half-pitch resolution of 1.1 μm.

Resolution improvement and phase retrieval

I quantified the performance of OFI-PSR algorithm by reconstructing a resolution test-target. **Figure 1.4**a shows the full FOV of the OFI-PSR reconstruction. Sample FOV is enlarged from 0.60 mm^2 to 4.56 mm^2, using a 0.35× demagnification camera adaptor. As a result, the effective pixel size at the focal plane of a 10×/0.3NA objective-lens is enlarged from 0.65 μm to 1.84 μm, which significantly downgrades the lateral resolution: an in-focus amplitude image of the sample is shown in **Figure 1.4**c, where the half-pitch resolution is ~2.2 μm. Using OFI-PSR algorithm with N = 5 out-of-focus measurements, I show in **Figure 1.4**d that the half-pitch resolution is improved to 1.1 μm, which also permits retrieval of the phase information of the sample as will be detailed below, increasing the overall SBP by a factor of ~8 (including both the amplitude and phase channels that are super-resolved). Note also that although the FOV with the camera adaptor

could in principle be 4.9 mm2, the geometrical mismatch between the circular output of the objective-lens and the rectangular sensor chip area causes a minor FOV loss at the corners, which results in an effective FOV of ~4.6 mm2, as illustrated in **Figure 1.5**a - also see **Table 1.1**.

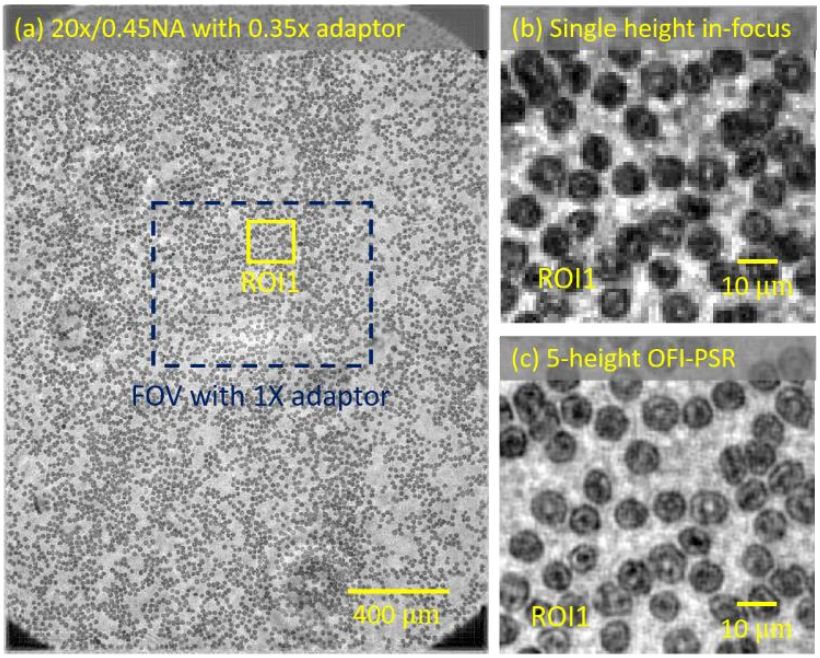

Figure 1.5 OFI-PSR reconstruction results for a human blood smear sample using N = 5 out-of-focus images. Microscope objective-lens: 20×/0.45NA; camera adapter: 0.35×; illumination wavelength: 532 nm. (a) Full FOV reconstruction of OFI-PSR algorithm. (b) In-focus image with a 0.35× camera adaptor shows undersampling. (c) OFI-PSR achieves a significant improvement in image quality.

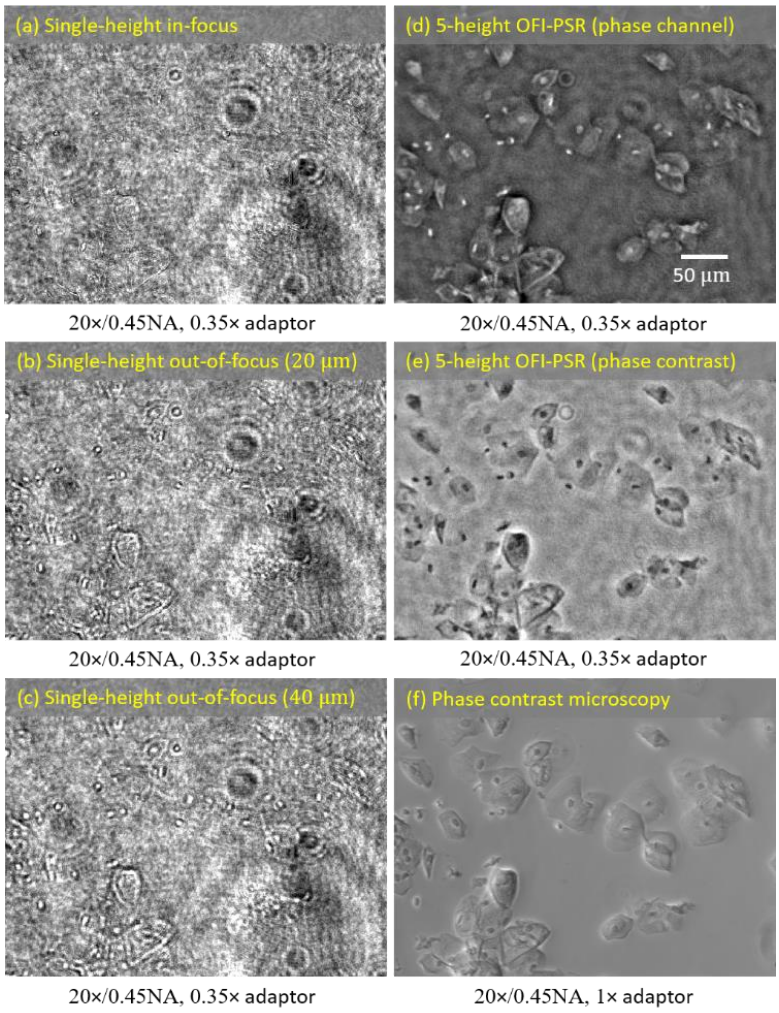

Figure 1.6 Demonstration of phase retrieval and pixel super-resolution by imaging an unstained Papanicolaou (Pap) smear. Microscope objective-lens: 20×/0.45NA; camera adapter: 0.35×; illumination wavelength: 532 nm. (a-c) In-focus and slightly out-of-focus intensity images of an unstained Pap smear sample. These images do not reveal much information about the sample since it is by and large a phase-only object. (d) The phase image recovered by OFI-PSR algorithm clearly reveals the structure and sub-cellular morphology of the cells. (e) Digital phase contrast image using the OFI-PSR reconstructed complex field. (f) Phase contrast image obtained using the same microscope with a 1× camera adaptor shows a good agreement with the digitally reconstructed phase image – except has ~8-fold smaller FOV compared to OFI-PSR.

After demonstrating the pixel super-resolution capabilities of computational out-of-focus imaging, next I imaged human blood cells and Pap smear samples (**Figure 1.5**, **Figure 1.6**). For

these experiments, I also used a 0.35× camera adaptor along with a 20×/0.45NA objective-lens, which resulted in a sample FOV of 1.14 mm^2. Using $N = 5$ out-of-focus intensity-only images, the full FOV reconstruction of a blood smear sample is shown in **Figure 1.5**a. A comparison of **Figure 1.5**b and c illustrates the significant improvement in image quality achieved with the OFI-PSR algorithm, restoring fine features of the sample from severely undersampled and out-of-focus image measurements.

In addition to pixel super-resolution, OFI-PSR also retrieves the object's phase information, and when combined with the amplitude channel, this increases the effective SBP by ~8-fold compared to an in-focus image of the object that shares the same FOV. **Figure 1.6**a-c show in-focus and slightly out-of-focus intensity images of an unstained Pap smear, which can be considered as a phase-only object since it is composed of a very thin layer of unstained cells taken from the cervix of a patient. That is why, the in-focus image in **Figure 1.6**a cannot reveal much information even if spatial undersampling were to be eliminated. However, as shown in **Figure 1.6**d, OFI-PSR recovers a high-resolution phase image of the sample, clearly revealing the structure and sub-cellular morphology of the cells. In **Figure 1.6**e, I also demonstrate a digital phase-contrast [50] image of the sample (calculated from **Figure 1.6**d), which provides a very good agreement with a phase contrast image obtained using the same microscope and a 1× camera adaptor, i.e., over an 8-fold smaller sample FOV compared to OFI-PSR.

Table 1.1 Comparison of OFI-PSR method with some of the traditional holographic imaging configurations. Experimental configuration: 10×/0.3NA objective-lens with a FN of 26.5 mm, QIClick CCD camera, 1392×1040 pixels, 6.45 μm pixel size and an illumination wavelength of 532 nm. The effective pixel count in this table considers both the amplitude and phase channels and assumes r = 2.

Configuration	Off-axis holography	Off-axis holography with lateral scanning	Two-step PSDH	Two-step PSDH with lateral scanning	OFI-PSR
	1× adaptor		0.35× adaptor		
FOV (mm^2)	0.60	12.22	0.60	3.05	4.56
Number of measurements	1	25-32	2	14-16	5
Half-pitch resolution (μm)	1.8	1.8	0.9	0.9	1.1
Effective pixel count (million)	0.37	7.54	1.48	7.54	7.54

Increased SBP and data efficiency of OFI-PSR

The SBP of a coherent computational imaging system is proportional to the number of effective pixels (N_I), reconstructed in a complex object image, i.e.,

$$N_I = 2 \cdot \mathrm{FOV} \cdot r^2 \cdot \left(\frac{\mathrm{NA}}{\lambda} \right)^2 , \tag{1.5}$$

where NA/λ is the cut-off spatial frequency of the coherent microscopy system, dictated by the NA of the objective-lens, r is the digital sampling factor along each direction (x and y), and the factor of 2 represents the independent spatial information contained in the phase and amplitude images of the complex sample. In my comparisons for different imaging configurations and coherent microscopy modalities, without loss of generality I assume $r = 2$. [51]

Based on these definitions, I compared the effective number of pixels and the data efficiency of my OFI-PSR method against conventional lateral-shift based FOV enhancement techniques for commonly used coherent imaging modalities (see **Table 1.1**). During these comparisons, to be fair

across different coherent imaging modalities, I utilized the same microscope objective-lens employed in my experiments (i.e., a 10×/0.3NA objective lens with a FN of 26.5 mm). Using N=5 out-of-focus intensity images, OFI-PSR can reconstruct a sample FOV of 4.56 mm^2 with an effective pixel count of $N_I = 7.54$ million. In my comparison, I first considered single-exposure off-axis holographic imaging configuration [52] which only keeps the real image quarter in the Fourier domain during the object reconstruction. Therefore, as summarized in **Table 1.1**, the lateral resolution is sacrificed compared to OFI-PSR and the effective pixel count of a single reconstructed complex object image is limited to 0.37 million. By using lateral scanning, in order for the off-axis holography based coherent microscopy to achieve the same SBP as in my method, an area of 12.22 mm^2 needs to be scanned. Note that to digitally stitch together several different FOVs, there is some spatial overlap that is required among images, which is typically ~10-20% on each side of the image. [53] This suggests that to cover 12.22 mm^2, 25-32 scanning positions and digital images are required, which is significantly larger compared to the number of out-of-focus images that OFI-PSR utilizes, i.e., 5.

Next I considered an alternative coherent imaging modality, i.e., the two-step phase-shifting digital holography (PSDH) configuration [54,55], which is expected to reach the diffraction limit of the imaging system using the least number of measurements among in-line holographic imaging configurations. Based on the use of the same objective-lens, a two-step PSDH configuration can reconstruct the image of a complex object with 2 measurements, achieving an effective pixel count of 1.48 million, which is four times larger than the off-axis holography configuration. To achieve the same SBP as in the OFI-PSR method, the two-step PSDH with lateral scanning would need to scan a FOV of 3.05 mm2. This means 7-8 scanning positions are needed in each phase-shifting

based imaging step, resulting in ~14-16 measurements in total, which is ~3× more than what OFI-PSR requires to achieve the same SBP, as also summarized in **Table 1.1**.

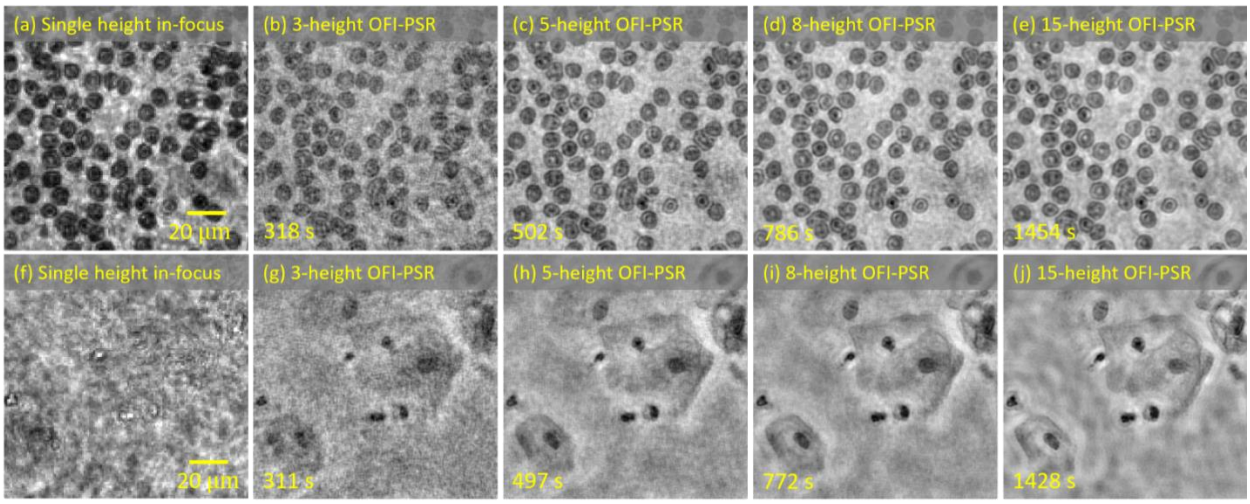

Figure 1.7 Reconstruction quality of OFI-PSR method with different number of out-of-focus measurements (N). Microscope objective-lens: 20×/0.45NA; camera adapter: 0.35×; illumination wavelength: 532 nm. I used an axial scanning step size of ~30 µm and each reconstruction has 100 iterations. The reconstruction time of each OFI-PSR image is shown at the left-bottom corner of each sub-figure. (a) In-focus undersampled image of a human blood smear sample. (b-e) OFI-PSR reconstructions of the same blood smear sample with N = 3, 5, 8 and 15 out-of-focus intensity images used as input. (f) In-focus undersampled image of an unstained Pap smear sample. (g-j) Digital phase contrast images of OFI-PSR reconstructions of the same Pap smear sample with N = 3, 5, 8 and 15 out-of-focus intensity measurements used as input.

Dependency of OFI-PSR reconstruction quality on the number of out-of-focus measurements

The quality of the reconstructed images using the OFI-PSR method is affected by the number of out-of-focus measurements, N. However, the required time for image acquisition and digital reconstruction increases linearly with the number of measurements, as also illustrated in **Figure 1.7**, where OFI-PSR based reconstructions of human blood cells and a Pap smear sample are compared using 3, 5, 8, and 15 different out-of-focus measurements, each with 100 iterations.

These results illustrate that, compared to the undersampled in-focus images (**Figure 1.7**a and f), OFI-PSR reconstructions with $N = 3$ measurements (**Figure 1.7**b and g) already show significantly improved features, although the aliasing signal is partially present. $N \geq 5$ further improves the reconstructed image quality and the high-frequency features are restored with good visibility. Since the data acquisition and digital computation time both increase linearly with the number of measurements, I conclude that N~5 out-of-focus measurements provides a good balance between imaging time and reconstruction quality.

Conclusion

I introduced a new computational out-of-focus imaging method termed OFI-PSR which helps to mitigate the SBP gap between microscope objective-lenses and opto-electronic image sensor chips to increase the SBP of coherent microscopy. I demonstrated the success of this wide-field imaging method using a conventional lens-based microscope and imaged resolution test-targets and biological samples. The OFI-PSR approach first extends the FOV of a single measurement using a demagnification camera adaptor, and then reconstructs a high-resolution complex image of the sample using an iterative algorithm. This super-resolution technique does not require lateral displacements between the specimen and the objective-lens, and also retrieves the phase information of the sample. To demonstrate the proof-of-concept of this approach, I used a 1.45-megapixel CCD camera and a 0.35× camera adaptor to achieve ~4.6 mm2 FOV with a 10×/0.3NA objective-lens, and mitigated undersampling-related artifacts using 5 out-of-focus intensity images, improving the SBP of the microscopic imaging system by a factor of ~8. Furthermore, OFI-PSR technique showed 3~6-fold reduction in the number of images required to achieve the same SBP using traditional in-line holography approaches. I believe this technique will broadly benefit

coherent imaging and holography fields and inspire new microscope designs with improved throughput and SBPs.

1.3 Deep learning techniques in microscopy

In this thesis, I will limit the discussion of deep learning to the scope of supervised learning, where a ground truth label is always provided for each image transformation or object detection task. For example, in the image super-resolution task, the DNN takes a low-resolution image as the input which was captured with a low-end microscope (e.g., with low-numerical aperture (NA) objectives), and performs a series of mathematical operations to the input image defined by the network structures. Then DNN output image, i.e., the network inference, is then compared against the ground truth image of the same field of view captured with a high-end microscope (e.g., with high-NA objectives) to calculate a loss function, e.g. mean square error. This loss function value indicates the performance of the network model in its current status and guides an optimizer function to fine tune the parameters inside the network model to satisfactory. Once the DNN model converges to a stead states, it can be used for inference task to new images that have never been seen by the network before. This forward inference is a single-pass operator without the need of iterative process, therefore can dramatically speed the image processing. Although the network training process is normally performed on graphic processing units (GPUs), the inference can be carried out on both GPUs and CPUs, and even embedded systems. Next, I will introduce the design and training of a DNN model.

1.3.1 Deep neural network architectures

A DNN model often consists of layers of different types that represent specific mathematical operations. Some common layers are described here:

Convolutional layer: an assembly of n (i.e., number of feature channels) convolution kernels, each performance a convolution operation of the input data:

$$x_{i+1} = w_{i,j} \otimes x_i + b_{i,j} \tag{1.6}$$

where x_i is the input from the previous layer, $w_{i,j}$ are the weights of the convolutional kernels, and $b_{i,j}$ is the bias term.

Fully-connect layer, or linear layer: connects all the neuron from the previous layer to the next layer through matrix multiplications.

Pooling layer: reduce the lateral dimensions of the input data by combining neighbouring neurons using the e.g., max value of a group of neurons.

Activation function: a nonlinear activation function that introduces non-linearity to a deep neural network. Rectified linear unit (ReLU) is one of the most used activation functions:

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1.7}$$

The layers are arranged according to a pre-defined network architecture. An example of network structures extensively used in my research is introduced in the next section.

1.3.2 **U-net**

One of the most popular network architecture for image processing and computer vision tasks is the U-net [56] whose variants are the architecture of the networks presented in Chapter 2 and Chapter 3. The original U-net (**Figure 1.8**) has a U shape that first contracts the input image and

increase the number of feature channels exponentially. Each contracting block consists of 2 convolutional layers, i.e.,

$$x_{k+1} = \text{Maxpool}(\text{ReLU}[\text{Conv}_{k2}\{\text{ReLU}[\text{Conv}_{k1}\{x_k\}]\}]) \tag{1.8}$$

where x_k is the input tensor to the *k*-th block and x_{k+1} is the corresponding output tensor.

Following 4 contracting blocks, the image is passed through a middle block which bridges the left contracting path and the right expansive path. Each expansive block also consists of 2 convolutional layers, arranges as:

$$y_k = \text{ReLU}[\text{Conv}_{k4}\{\text{ReLU}[\text{Conv}_{k3}\{\text{Cat}[x_{k+1}, \text{Upsample}(y_{k+1})]\}]\}] \tag{1.9}$$

where the CAT[.] operator represent the concatenation operation in the feature channel dimensions, the upsample(.) operator represents up-convolutional layer that resizes the input image to two time big. The U-net structure allow a neural network model to learn the image transformation at different scales and feature levels, and routinely achieves satisfactory results in our research.

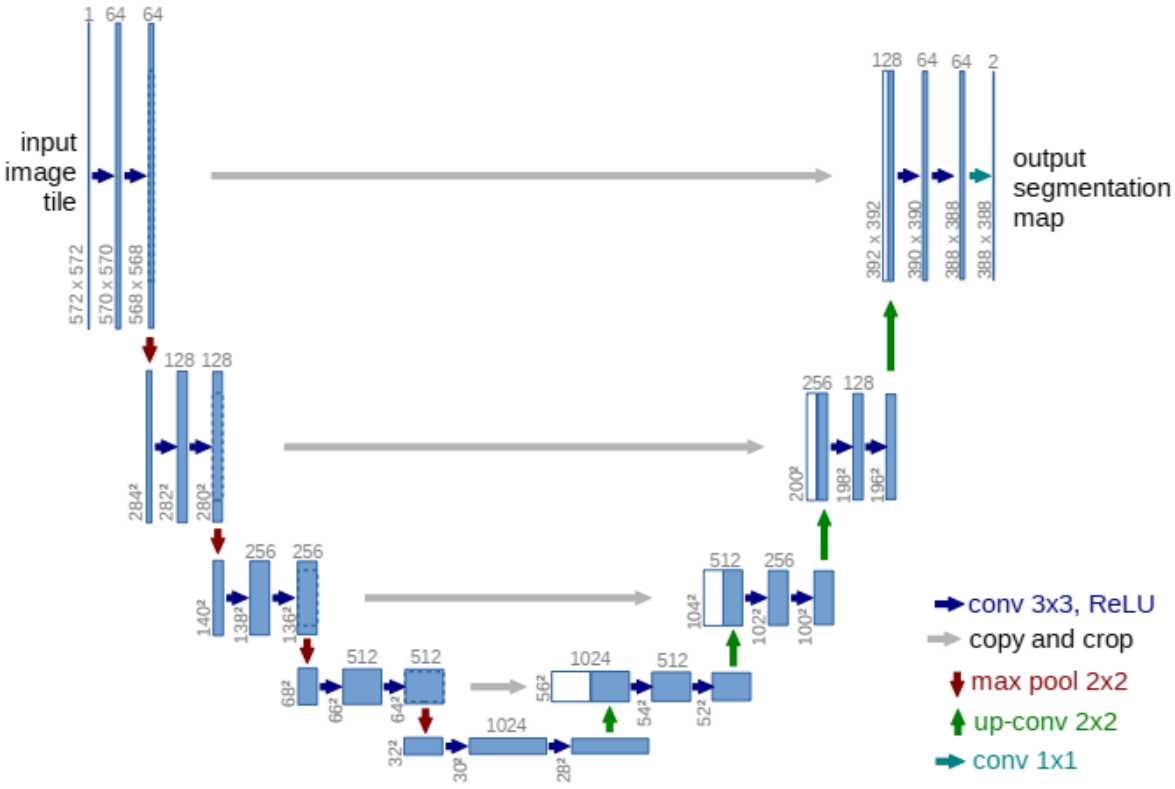

Figure 1.8 The original U-net architecture. [56]

1.3.3 **Generative adversarial networks (GANs)**

Other than training a single network model, one can also train multiple network models that are assembled under a single framework, such as the generative adversarial networks (GANs). [57] In the GAN framework, there are two networks: one is the generator network *G* that transforms the input data into the form we desire. The other one is the discriminator network *D* that learns to discriminate the generator inferred images and the ground truth images and attribute its discrimination result to (part of) the generator loss. Two optimizer functions are employed to simultaneously optimize the two networks. Specifically, the generator's optimizer minimizes the generator loss function:

$$\mathcal{L}(G;D) = -\log D\big(G\big(x\big)\big) \tag{1.10}$$

The Discriminator's optimizer minimizes the discriminator loss function:

$$\mathcal{L}(D;G) = -\log D\left(y\right) - \log\left[1 - D\left(G\left(x\right)\right)\right] \tag{1.11}$$

where x is the input data, y is the ground truth image, and $G(x)$ is the generator inferred image for input x. In my implementation for cross-modality image transformation tasks, the generator loss also has additional pixel-wise loss so that the generator network output is conditional to the input image, e.g., in case mean square error (MSE) loss is used:

$$\mathcal{L}(G;D) = -\log D\left(G\left(x\right)\right) + \nu \times \mathrm{MSE}(G(x), y) \tag{1.12}$$

where ν is a constant weight value that defines the ratio of pixel-wise loss in the total generator loss. Once the training is stabilized, the two networks are in an equilibrium status, meaning the generator keeps improving the output image quality and trying to fool the discriminator, while the discriminator keeps improving discrimination capabilities and looking for even the smallest artefacts and feeds them back to the generator network.

1.3.4 **Image pre-processing**

The performance of a cross-modality image transformation network that are trained with supervised learning heavily relies on high quality training image pairs. Although the network input and the ground truth images can be synthesized according to a specific model of the imaging systems, it is not an ideal approach since the physical assumptions cannot be perfectly accurate due to the variations in hardware, experiment environments, and even sample preparation protocols. In my research the input and ground truth images are mostly experimentally captured with different imaging setups, e.g., using different objective lenses or imaging modalities. Consequently, the two sets of images need to be cross-registered to sub-pixel level accuracy for

the optimizers to minimize the pixel-wise losses. One of the difficulties of image cross-registration comes from the chromatic and spectral aberrations of different optical systems. Such aberrations are non-uniform across the FOV of an image, therefore, require an elastic registration process. In practice, the image registration processes are often carried out in several steps with different registration scales, designed for the experimental datasets, as described in sections 1.4, 2.6. and 3.5. One special case for image cross-registration is described in 3.5, where the input images (autofluorescence) and ground truth images (bright-field) are vastly different, therefore hard to be registered by directly calculating the cross-correlation map. In such cases, one can train a "registration network" with roughly registered dataset, i.e., with ~10-pixel lateral shifts between the input and the ground truth images. The output images from this "registration network" are often non-ideal, but they can help calculate local shifts and reduce the lateral mismatches. Sometimes several registration network models need to be trained repeatedly, during which time the training dataset gradually converges to a well-registered state.

1.3.5 **Neural network model training and validation**

Once the dataset is precisely registered, it is then divided into the training and the validation datasets. The training dataset is for the network to learning the transformation, and the validation dataset is to monitor the performance of the network model but avoids giving any feedback to the optimizer. During the training process, the parameters of a neural network model are optimized by an optimizer function, e.g. Adam optimizer [58], through iterative forward propagation of the input images and backward propagation of the loss values. At the end of each training epoch (i.e., all the input images have gone through a forward propagation), a loss value is calculated using the validation dataset and the best network model with the lowest validation loss is selected. The

training method and schedule will be described in the methods section for each work to be demonstrated.

Here I will first starting with a project that employs a deep network to learn the statistical transformations between the mobile and optimized benchtop microscope images to create a convoluted mapping between the two imaging instruments, which includes not only a spatially and spectrally varying distorted point-spread function and the associated color aberrations, but also a non-uniform space warping at the image plane, introduced by the inexpensive mobile-phone microscope. Unlike most image enhancement methods, this work does not consider physical degradation models during the image formation process. Such image degradation models are in general hard to estimate theoretically or numerically, which limits the applicability of standard inverse imaging techniques. Moreover, even if such a forward model could be estimated, there are almost always unknown and random deviations from it due to fabrication tolerances and alignment imperfections that are unfortunately unavoidable in large scale manufacturing. Instead of trying to come up with such a forward model for image degradation, the deep neural network learns how to predict the benchtop microscope image that is most statistically likely to correspond to the input smartphone microscope image by learning from experimentally-acquired training images of different samples.

1.4 Deep learning enhanced mobile-phone microscopy

Introduction

Optical imaging is a ubiquitous tool for medical diagnosis of numerous conditions and diseases. However, most of the imaging data, considered the gold standard for diagnostic and screening purposes, are acquired using high-end benchtop microscopes. Such microscopes are

often equipped with expensive objectives lenses and sensitive sensors, are typically bulky, must be operated by trained personnel, and require substantial supporting infrastructure. These factors potentially limit the accessibility of advanced imaging technologies, especially in resource-limited settings. Consequently, in recent years researchers have implemented cost-effective, mobile microscopes, which are often based on off-the-shelf consumer electronic devices, such as smartphones and tablets [59]. As a result of these research efforts, mobile-phone-based microscopy has demonstrated promise as an analytical tool for rapid and sensitive detection and automated quantification of various biological analytes as well as for the imaging of, e.g., pathology slides [59–69].

An important challenge in creating high-quality benchtop microscope equivalent images on mobile devices stems from the motivation to keep mobile microscopes cost-effective, compact and light-weight. Consequently, most mobile microscope designs employ inexpensive, often battery-powered illumination sources, such as light-emitting diodes (LEDs), which introduce color distortions into the acquired images. Furthermore, mobile microscopes are usually equipped with low numerical apertures (NAs) also containing aberrated and often misaligned optical components, which add further distortions into the acquired images at the micro-scale. Although the lenses of mobile-phone cameras have advanced significantly over the last several years, large volume fabrication techniques are employed in the moulding and assembly of these plastic lenses, which creates random deviations for each mobile camera unit compared with the ideal optical design and alignment. Some of these distortions also vary to some degree as a function of time and usage, due to, e.g., the battery status of the mobile device and the illumination unit, the poor mechanical alignment precision of the sample holder, and the user experience. Furthermore, since most optoelectronic imagers found in consumer electronic devices including smartphones have been

optimized for close and mid-range photography rather than microscopy, they also contain built-in design features such as varying micro-lens positions with respect to the pixels, which create additional spatial and spectral distortions for microscopic imaging. Finally, since mobile-phone cameras have small pixel sizes (on the order of 1-2 μm) with a very limited capacity of a few thousand photons per pixel, such mobile imagers also have reduced sensitivity. In contrast, high-end benchtop microscopes that are used in medical diagnostics and clinical applications are built around optimized illumination and optical pick-up systems with calibrated spectral responses, including diffraction-limited and aberration-corrected objective lenses and highly-sensitive CCDs (charged-coupled devices) with large pixels.

Here, I describe the substantial enhancement of the imaging performance of a bright-field mobile-phone based microscope using deep learning. The mobile microscope was implemented using a smartphone with a 3D-printed optomechanical attachment to its camera interface, and the image enhancement and color aberration correction were performed computationally using a deep convolutional neural network (**Figure 1.9**). Deep learning [70] is a powerful machine learning technique that can perform complex operations using a multi-layered artificial neural network and has shown great success in various tasks for which data are abundant [71–74]. The use of deep learning has also been demonstrated in numerous biomedical applications, such as diagnosis [75,76], image classification [77], among others [12,78–81]. In the presented method, a supervised learning approach is first applied by feeding the designed deep network with input (smartphone microscope images) and labels (gold standard benchtop microscope images obtained for the same samples) and optimizing a cost function that guides the network to learn the statistical transformation between the input and label. Following the deep network training phase, the network remains fixed and a smartphone microscope image input into the deep network is rapidly

enhanced in terms of spatial resolution, signal-to-noise ratio, and color response, attempting to match the overall image quality and the field of view (FOV) that would result from using a 20× objective lens on a high-end benchtop microscope. In addition, the image output by the network will be demonstrated to have a larger depth of field (DOF) than the corresponding image acquired using a high-NA objective lens on a benchtop microscope. Each enhanced image of the mobile microscope is inferred by the deep network in a non-iterative, feed-forward manner. For example, the deep network generates an enhanced output image with a FOV of ~0.57 mm^2 (the same as that of a 20× objective lens), from a smartphone microscope image within ~0.42 s, using a standard personal computer equipped with a dual graphics-processing unit. This deep learning-enabled enhancement is maintained even for highly compressed raw images of the mobile-phone microscope, which is especially desirable for storage, transmission and sharing of the acquired microscopic images for e.g., telemedicine applications, where the neural network can rapidly operate at the location of the remote professional who is tasked with the microscopic inspection of the specimens.

Materials and methods

Design of the Smartphone-Based Microscope:

A Nokia Lumia 1020 smartphone was used in the design of the smartphone-based transmission microscope. It has a CMOS image sensor chip with an active area of 8.64 mm × 6 mm, and a pixel size of 1.12 μm. The built-in camera of the smartphone is formed with 6-lenses, a combination of one glass lens (facing the prototype) and five additional plastic lenses. The smartphone sensor aperture is f/2.2 [82].The regular camera application of the smartphone facilitates the capture of images in raw format (i.e., DNG) as well as JPG images using the rear camera of the smartphone, which has 41 megapixels. The same application also provides

adjustable parameters such as the sensor's sensitivity (International Organization for Standardization, ISO) and exposure time. While capturing images, the operator set the ISO to 100, exposure time and focus to auto, and white balance to cloud mode, which is a predefined mode that had been visually evaluated as one of the best modes for imaging pathology slides. The automatically adjusted exposure times for the smartphone microscope images ranged from 1/49 to 1/13 s.

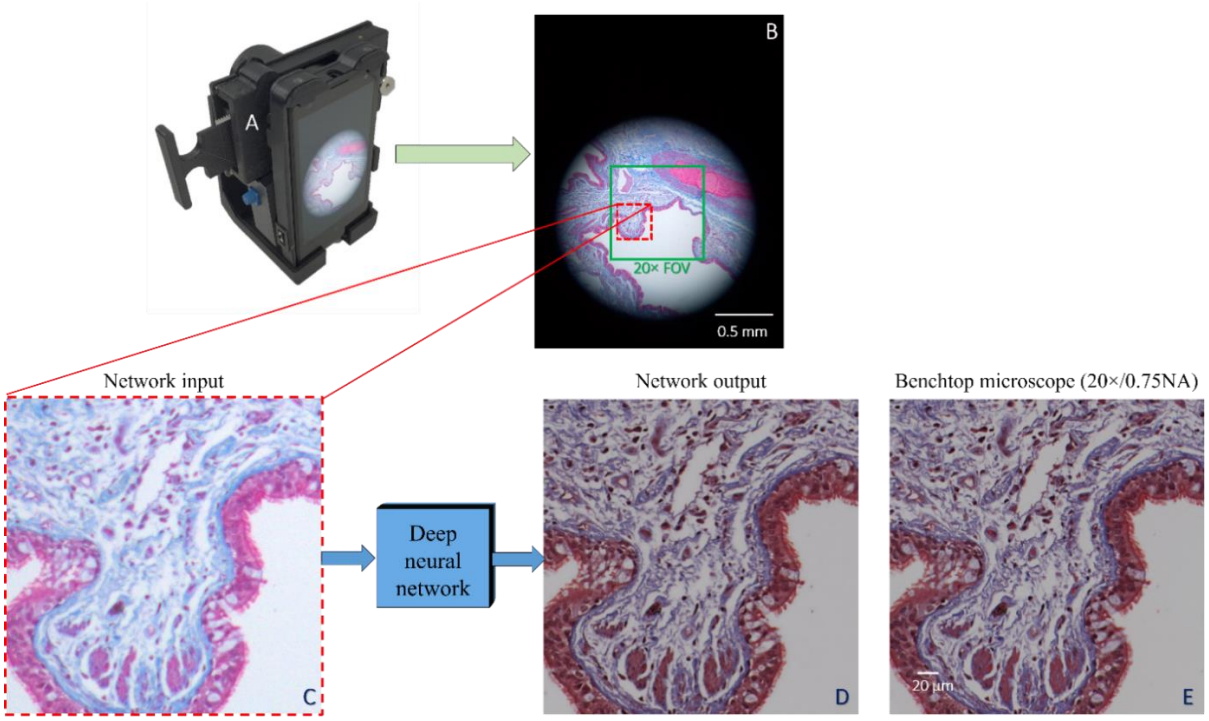

Figure 1.9 Deep learning enhanced mobile-phone microscopy. (**A**, **B**) Masson's-trichrome-stained lung tissue sample image acquisition using a cost-effective smartphone microscope device. (**C**) Input region of interest (ROI), for which the deep network blindly yields (**D**) an improved output image, which resembles (**E**) an image obtained using a high-end benchtop microscope, equipped with a 20×/0.75NA objective lens and a 0.55NA condenser.

Autodesk Inventor was used to design the 3D layout of the optomechanical attachment unit that transforms the smartphone into a field-portable and cost-effective microscope. It includes an *xyz* stage that facilitates lateral scanning and axial focusing. The optomechanical parts of the unit

were printed using a 3D printer (Stratasys, Dimension Elite) and acrylonitrile butadiene styrene (ABS).

To provide bright-field illumination, a 12 RGB LED ring structure (NeoPixel Ring) with integrated drivers (product no. 1643) and its microcontroller (product no. 1501) were purchased from Adafruit (New York City, NY, USA). The LEDs in the ring were programmed using Arduino to provide white light to illuminate the samples. The LEDs were powered using a rechargeable battery (product no. B00EVVDZYM, Amazon, Seattle, WA, USA). The illumination unit illuminated each sample from the back side through a polymer diffuser (Zenith Polymer® diffuser, 50% transmission, 100 μm thickness, product no. SG 3201, American Optic Supply, Golden, CO, USA). An external lens with a focal length of 2.6 mm, provided a magnification of ~2.77, a FOV of ~1 mm^2, and a half-pitch lateral resolution of ~0.87 μm. The *xy* stage on the sample tray was used to move each sample slide for lateral scanning and the *z* stage to adjust the depth of focus of the image.

Benchtop Microscope Imaging:

Gold standard image data acquisition was performed using an Olympus IX83 microscope equipped with a motorized stage. The images were acquired using a set of Super Apochromat objectives, (Olympus UPLSAPO 20X/0.75NA, WD0.65). The color images were obtained using a Qimaging Retiga 4000R camera with a pixel size of 7.4 μm. The microscope was controlled by MetaMorph® microscope automation software (Molecular Devices, LLC), which includes automatic slide scanning with autofocusing. The samples were illuminated using a 0.55NA condenser (Olympus IX2-LWUCD).

Sample Preparation:

All the human samples were obtained after de-identification of the patients and related information and were prepared from existing specimens. Therefore, this work did not interfere with the standard care practices or sample collection procedures.

Lung tissue: De-identified formalin-fixed paraffin-embedded Masson's-trichrome-stained lung tissue sections from two patients were obtained from the Translational Pathology Core Laboratory at UCLA. The samples were stained at the Histology Lab at UCLA.

Pap smear: A de-identified Pap smear slide was provided by UCLA Department of Pathology.

Blood smear: A de-identified human blood smear slide was provided by UCLA Microbiology Lab.

Data Preprocessing:

To ensure that the deep network learns to enhance smartphone microscope images, it is important to pre-process the training image data so that the smartphone and benchtop microscope images will match. The deep network learns how to enhance the images by following an accurate smartphone and benchtop microscope FOV matching process, which in this designed network is based on a series of spatial operators (convolution kernels). Providing the deep network with accurately registered training image data enables the network to focus the learning process on correcting for repeated patterns of distortions between the images (input vs. gold standard), making the network more compact and resilient overall and requiring less data and time for training and data inference.

This image registration task is divided into two parts. The first part matches the FOV of an image acquired using the smartphone microscope with that of an image captured using the benchtop microscope. This FOV matching procedure can be described as follows: (i) Each cell phone image is converted from DNG format into TIFF (or JPEG) format with the central 0.685

mm^2 FOV being cropped into four parts, each with 1024×1024 pixels. (ii) Large-FOV, high-resolution benchtop microscope images (~25K×25K pixels) are formed by stitching 2048×2048-pixel benchtop microscope images. (iii) These large-FOV images and the smartphone image are used as inputs for scale-invariant feature transform (SIFT) [83] and random sample consensus (RANSAC) algorithms. First, both color images are converted into grey-scale images. Then, the SIFT frames (*F*) and SIFT descriptors (*D*) of the two images are computed. *F* is a feature frame and contains the fractional centre of the frame, scale, and orientation. *D* is the descriptor of the corresponding frame in *F*. The two sets of SIFT descriptors are then matched to determine the index of the best match. (iv) A homography matrix, computed using RANSAC, is used to project the low-resolution smartphone image to match the FOV of the high-resolution benchtop microscope image, used as gold standard.

Following this FOV matching procedure, the smartphone and benchtop microscope images are globally matched. However, they are not accurately registered, mainly due to distortions caused by the imperfections of the optical components used in the smartphone microscope design and inaccuracies originating during the mechanical scanning of the sample slide using the *xyz* translation stage. This second part of the registration process locally corrects for all these distortions between the input and gold standard images by applying a pyramid elastic registration algorithm, which is depicted in **Figure 1.10**. During each iteration of this algorithm, both the smartphone and corresponding benchtop microscope images are divided into $N \times N$ blocks, where typically $N = 5$. A block-wise cross-correlation is calculated using the corresponding blocks from the two images. The peak location inside each block represents the shift of its centre. The peak value, i.e., the Pearson correlation coefficient [84], represents the similarity of the two blocks. A cross-correlation map (CCM) and an $N \times N$ similarity map are extracted by locating the peak

locations and fitting their values. An $m \times n$ translation map is then generated based on the weighted average of the CCM at each pixel. This translation map defines a linear transform from the distorted image to the target enhanced image. This translation operation, although it corrects distortions to a certain degree, is synthesized from the block-averaged CCM and therefore should be refined with smaller-block-size CCMs. In the next iteration, N is increased from 5 to 7, and the block size is reduced. This iterative procedure is repeated until the minimum block size is reached, which was empirically set to be $m \times n = 50 \times 50$ pixels. The elastic registration in each loop followed the open-source NanoJ plugin in ImageJ [85,86].

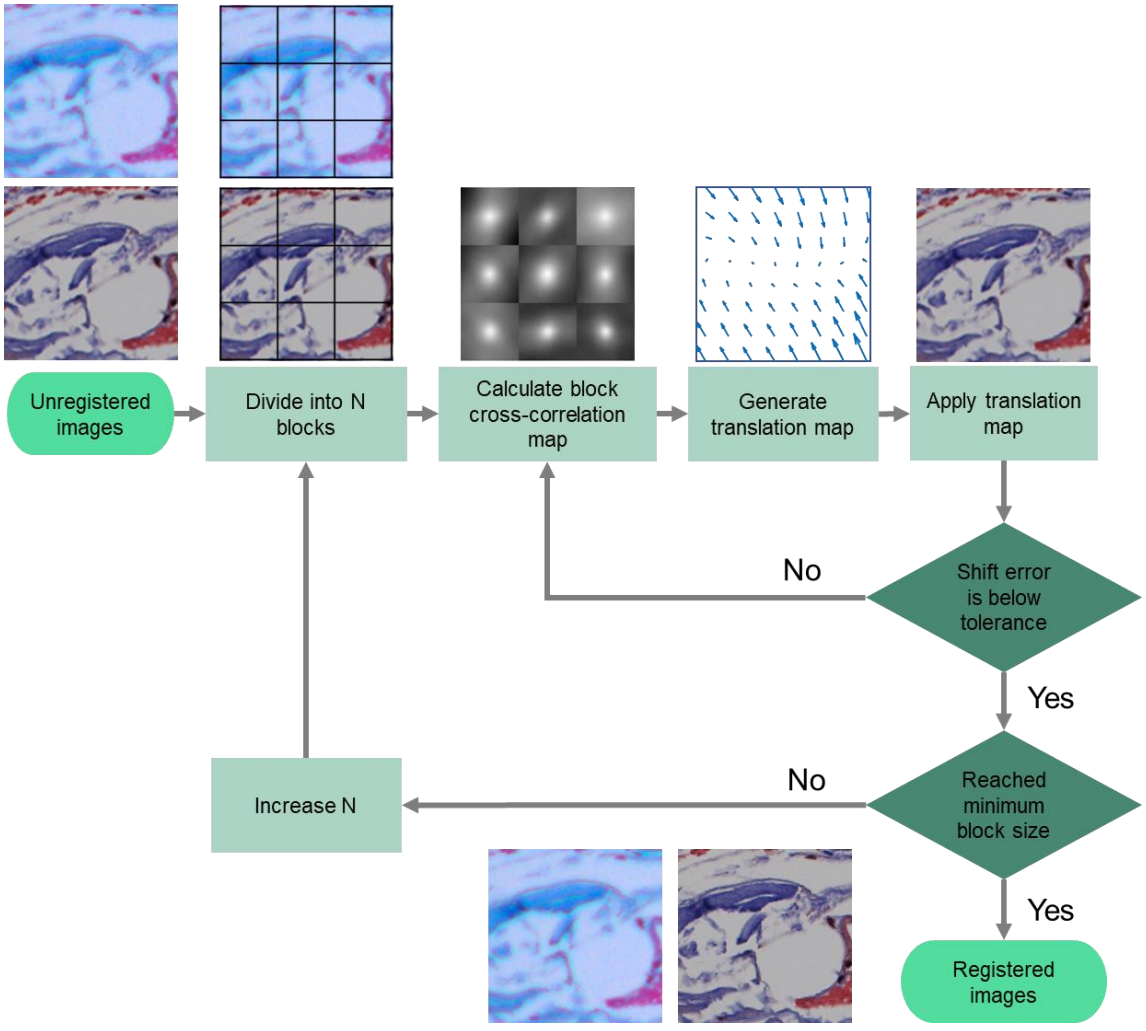

Figure 1.10 Pyramid elastic registration algorithm

Following the FOV matching and registration steps discussed above, the last step is to upsample the target image in a way that will enable the network to learn the statistical transformation from the low-resolution smartphone images into high-resolution, benchtop-microscope equivalent images. When the benchtop microscope was used to create gold standard images used for training, each sample was illuminated using a 0.55NA condenser, which creates a theoretical resolution limit of approximately 0.4 μm using a 0.75 NA objective lens (20×). However, the lateral resolution is constrained by the effective pixel size at the CCD, which is 7.4 μm; therefore, the practical half-pitch resolution of the benchtop microscope using a 20× objective lens is: 7.4 μm / 20 = 0.37 μm, corresponding to a period of 0.74 μm. On the other hand, the smartphone microscope is based on a CMOS imager and has a half-pitch resolution of 0.87 μm, corresponding to a resolvable period of 1.74 μm. Thus, the desired upsampling ratio between the smartphone and benchtop microscope images is given by 0.87 / 0.37 = 2.35. Therefore, the deep network was trained to upsample by a ratio of 2.5, and by applying the upsampling only at the final convolutional layers, the network structure was enabled to remain compact, making it easier to train and infer [87].

Deep Neural Network Architecture and Implementation:

The deep neural network architecture [79] receives three input feature maps (RGB channels), and following the first convolutional layer, the number of feature maps is expanded to 32. Formally, the convolution operator of the *i*-th convolutional layer for *x,y*-th pixel in the *j*-th feature map is given by:

$$g_{i,j}^{x,y} = \sum_{r} \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} w_{i,j,r}^{u,v} g_{i-1,r}^{x+u,y+v} + b_{i,j} \tag{1.13}$$

where g defines the feature maps (input and output), $b_{i,j}$ is a learned bias term, r is the index of the feature maps in the convolutional layer, and $w_{i,j,r}^{u,v}$ is the learned convolution kernel value at its u,v-th entry. The size of the convolutional kernel is $U \times V$, which was set to be 3×3 throughout the network. Following the initial expansion of the number of feature maps from 3 to 32, the network consists of five residual blocks, which contribute to the improved training and convergence speed of the deep networks [71]. The residual blocks implement the following structure:

$$X_{k+1} = X_k + \mathrm{ReLU}(\mathrm{Conv}_{k\_2}(\mathrm{ReLU}(\mathrm{Conv}_{k\_1}(X_k)))) \tag{1.14}$$

where $\mathrm{Conv}(.)$ is the operator of each convolutional layer, and the non-linear activation function that was applied throughout the deep network was ReLU, defined as $\mathrm{ReLu}(x) = \max(0, x)$. The number of feature maps for the k-th residual block is given by [88]

$$A_k = A_{k-1} + \mathrm{floor}((\alpha \times k) / K + 0.5) \tag{1.15}$$

where $K = 5$ is the total number of residual blocks, $k = [1:5]$, $\alpha = 10$, and $A_0 = 32$. By gradually increasing the number of feature maps throughout the deep network (instead of having a constant large number of feature maps), the network was kept more compact and less demanding on computational resources (for both training and inference). However, increasing the number of channels through residual connections creates a dimensional mismatch between the features represented by X_k and X_{k+1} in equation(1.14). To avoid this issue, X_k was augmented with zero-valued feature maps, to match the total number of feature maps in X_{k+1}. Following the output of the fifth residual block, another convolutional layer increases the number of feature maps from 62 to 75. The following two layers transform these 75 feature maps, each with $S \times T$ pixels, into three output channels, each with $(S \times L) \times (T \times L)$ pixels, which correspond to the RGB channels of the

target image. In this case, L was set to 2.5 (as detailed in the Data Preprocessing section, related to upsampling). To summarize, the number of feature maps in the convolutional layers in the deep network follows the sequence of: 3 → 32 → 32 → 34 → 34 → 38 → 38 → 44 → 44 → 52 → 52 → 62 → 62 → 75 → 3 → 3. If the number of pixels in the input is odd, the size of the output is given by $3\times\lceil(S\times L)\rceil\times\lceil(T\times L)\rceil$. Performing upsampling only at the final layers further reduces the computational complexity, increases the training and inference speed, and enables the deep network to learn an optimal upsampling operator.

The network was trained to optimize the cost function ℓ based on the current network output $Y^{\Theta}=\Phi(X_{input};\Theta)$ and the target (benchtop microscope) image Y^{Label}:

$$\ell(\Theta)=\frac{1}{3\times S\times T\times L^2}\left[\sum_{c=1}^{3}\sum_{s=1}^{S\times L}\sum_{t=1}^{T\times L}\left\|Y_{c,s,t}^{\Theta}-Y_{c,s,t}^{Label}\right\|_2^2+\lambda\sum_{c=1}^{3}\sum_{s=1}^{S\times L}\sum_{t=1}^{T\times L}\left|\nabla Y^{\Theta}\right|_{c,s,t}^2\right], \tag{1.16}$$

where X_{input} is the network input (smartphone microscope raw image), with the deep network operator denoted as Φ and the trainable network parameter space as Θ. The indices *c*, *s*, and *t* denote the *s*,*t*-th pixel of the *c*-th color channel. The cost function (equation (1.16)) balances the mean-squared error and image sharpness with a regularization parameter λ, which was set to be 0.001. The sharpness term, $\left|\nabla Y^{\Theta}\right|_{c,s,t}^2$ is defined as [89] $\left|\nabla Y^{\Theta}\right|^2=\left(h*Y^{\Theta}\right)^2+\left(h^T*Y^{\Theta}\right)^2$, where

$$h=\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}, \tag{1.17}$$

and $(.)^T$ is the matrix transpose operator.

The calculated cost function is then back-propagated to update the network parameters (Θ), by applying the adaptive moment estimation optimizer (Adam) [58] with a constant learning rate

of 2×10^{-4}. During the training stage, the network was trained with a mini-batch of 32 patches (**Table 1.2**). The convolution kernels were initialized by using a truncated normal distribution with a standard deviation of 0.05 and a mean of 0 [71]. All the network biases were initialized as 0.

Table 1.2 Deep neural network training details for different samples. All the images were captured using the smartphone automatic exposure settings.

	Number of input–output patches (number of pixels in each mobile phone microscope image)	Validation set (number of pixels in each mobile phone microscope image)	Number of epochs till convergence	Training time
Masson's trichrome stained lung tissue	129,472 patches (60×60 pixels)	95 images (800×800 pixels)	134	36 h, 40 min
H&E stained Pap smear	222,008 patches (60×60 pixels)	63 images (1024×1024 pixels)	190	20 h, 24 min
Blood Smear	65,520 patches (60×60 pixels)	9 images (1024×1024 pixels)	206	10 h, 25 min

Color distance calculations:

The CIE-94 color distance was developed by the Commission internationale de l′éclairage (CIE) [90], [91], and was used it as a metric to quantify the reconstruction quality of the deep network, with respect to the gold standard benchtop microscope images of the same samples. The average and the standard deviation of the CIE-94 were calculated between the 2.5× bicubic upsampled smartphone microscope raw *input* images and the benchtop microscope images (used as gold standard), as well as between the deep network *output* images and the corresponding benchtop microscope images, on a pixel-by-pixel basis and averaged across the images of different samples (

Table 1.3). As reported in

Table 1.3, the CIE-94 color difference calculations [90] were also performed on warp-corrected (using the pyramid elastic registration algorithm) and 2.5× bicubic upsampled smartphone microscope images as well as on their corresponding network output images, all calculated with respect to the same gold standard benchtop microscope images.

Results and discussion

Schematics of the deep network training process are shown in **Figure 1.11**. Following the acquisition and registration of the smartphone and benchtop microscope images (see the Data Preprocessing subsection in Materials and Methods section), where the benchtop microscope was equipped with a 20× objective lens (NA = 0.75), the images were partitioned into input and corresponding label pairs. Then, a localized registration between input and label was performed using pyramid elastic registration to correct distortions caused by various aberrations and warping in the input smartphone microscope images (see 0 Materials and methods, **Figure 1.10**, and **Figure 1.11**). These distortion-corrected images were divided into training and validation sets. An independent testing set (which was *not* aberration-corrected) enabled us to blindly test the network on samples that were not used for the network training or validation.

The training dataset was generated by partitioning the registered images into 60 × 60 pixel and 150 × 150 pixel patch images (with 40% overlap), from the distorted smartphone and the gold standard benchtop microscope images, respectively (the numbers of training patches and the required training times for the different samples are provided in **Table 1.2**). Multiple networks were trained corresponding to multiple types of pathology samples such as stained lung tissue, Papanicolaou (Pap) and blood smear samples, while maintaining the exact same neural network architecture. Following the training of the deep networks (**Table 1.2**), the networks remained fixed and were used to blindly test samples from different pathology slides.

Table 1.3 Average and standard deviation (Std) of the CIE-94 color distances compared to the gold standard benchtop microscope images for the different pathology samples.

	(A) Raw smartphone microscope images		(B) Warp-corrected smartphone microscope images		(C) Deep network output images of (A)		(D) Deep network output images of (B)	
	Average	Std	Average	Std	Average	Std	Average	Std
Masson's-trichrome-stained lung tissue (TIFF)	15.976	1.709	16.077	1.683	4.369	0.917	3.814	0.797
Masson's-trichrome-stained lung tissue (JPEG)	15.914	1.722	15.063	1.820	4.372	0.847	3.747	0.908
H&E-stained Pap smear (TIFF)	26.230	0.766	23.725	0.969	2.127	0.267	2.092	0.317
Blood smear (TIFF)	20.645	0.795	20.601	0.792	1.816	0.115	1.373	0.052

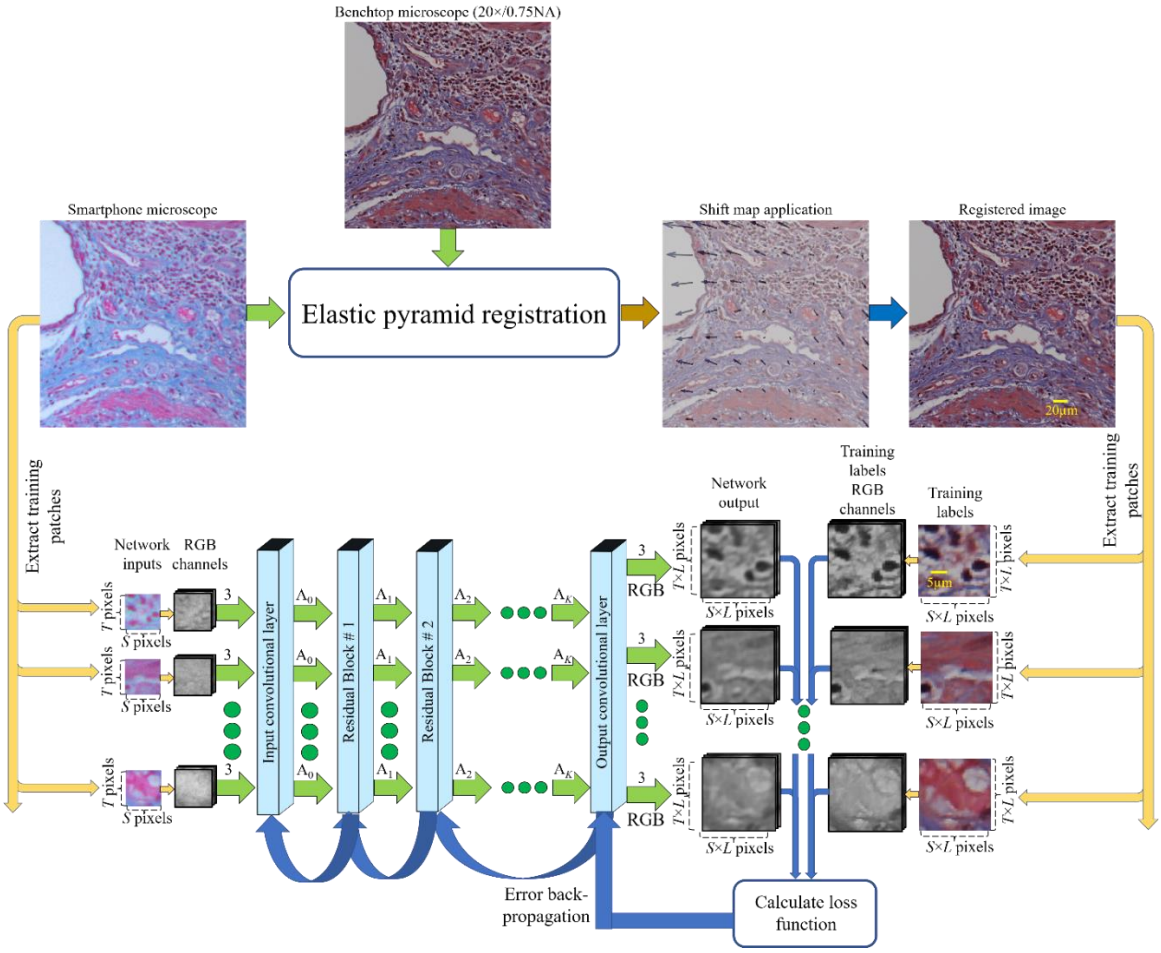

Figure 1.11 Training phase of the deep neural network

First, the deep learning framework was applied to Masson's-trichrome-stained lung tissue. A representative result is shown in **Figure 1.12**, which demonstrates the ability of the deep network to restore spatial features that cannot be detected in the raw smartphone microscope image due to various factors including spatial blurring, poor signal-to-noise ratio, non-ideal illumination, and the spectral response of the sensor. Following the inference of the deep network acting on the input smartphone microscope image, several spatial details were restored as illustrated **Figure 1.12**D and G. In addition, the deep network corrected the severe color distortion of the smartphone image, restoring the original colors of the dyes that were used to stain the lung tissue sample, which is highly important for telepathology and related applications. As detailed in

Table 1.3 and 0 Materials and methods, the CIE-94 color distance [90] was used as a metric to quantify the reconstruction quality of the deep network, with respect to the gold standard benchtop microscope images of the same samples. Overall, the deep network has significantly improved the average CIE-94 color distance of the mobile microscope images by a factor of 4~11 fold, where the improvement was sample dependent as shown in

Table 1.3. This color improvement is especially significant for pathology field, where different dyes are used to stain various tissue structures, containing critical information for expert diagnosticians. Another advantage of applying the deep network is the fact that it performs denoising of the smartphone microscope images, while retaining the fidelity of the fine-resolution features, as demonstrated in **Figure 1.12**(I1,I2,I3). These results were also quantitatively evaluated by using the structural similarity (SSIM) index [92] calculated against the gold standard images, revealing the improvement of the neural network output images as shown in **Table 1.4**.

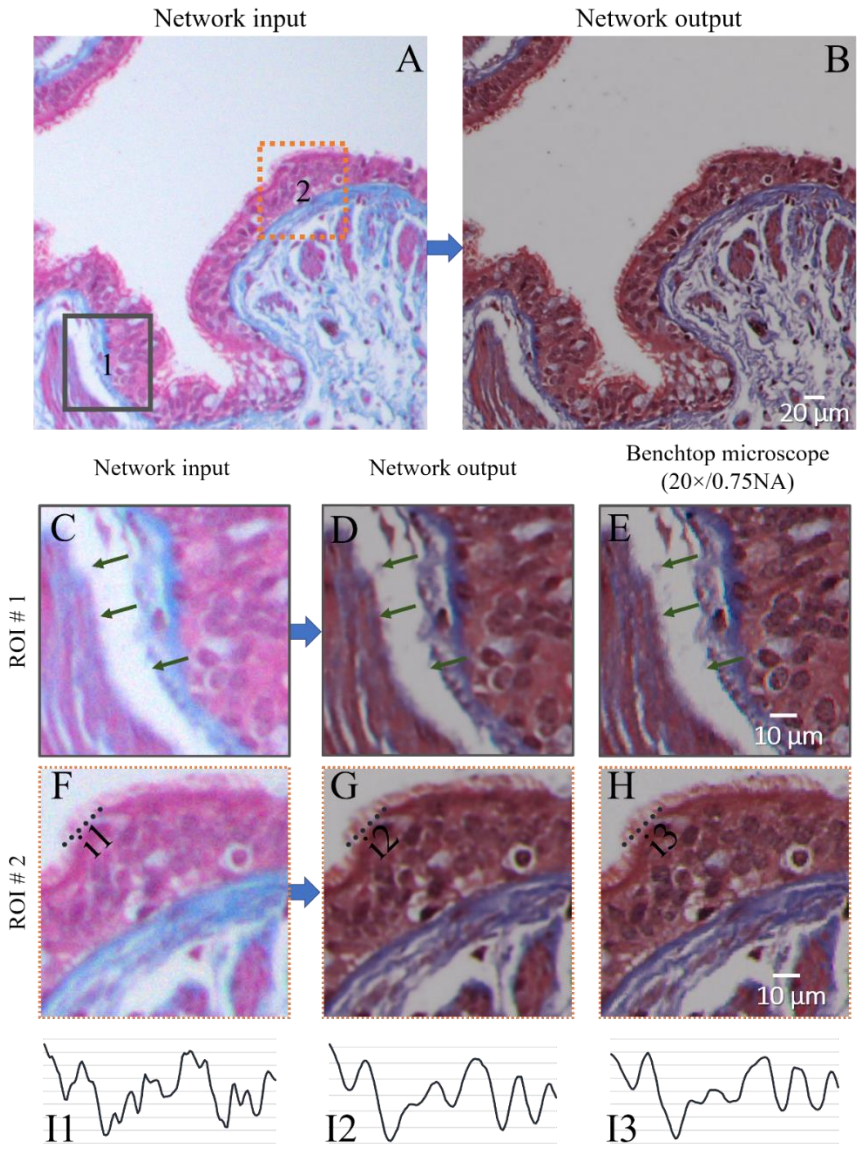

Figure 1.12 Deep neural network output for a Masson's-trichrome-stained lung tissue section. (**A**) Smartphone microscope image, and (**B**) its corresponding deep network output. Zoomed-in versions of the ROIs shown in (**C**, **F**) the smartphone input image and (**D**, **G**) the neural network output image. (**E**, **H**) Images of the same ROIs acquired using a 20×/0.75NA objective lens (with a 0.55NA condenser). The green arrows in (**C**, **D**, **E**) point to some examples of the fine structural details that were recovered using the deep network. Several other examples can be found in (**D**, **G**) compared to (**C**, **F**), which altogether highlight the significant improvements in the deep network output images, revealing the fine spatial and spectral details of the sample. (**I**) Cross-section line profiles from (**F**, **G**, **H**) demonstrating the noise removal performed by the deep network, while retaining the high-resolution spatial features.

Table 1.4 Average SSIM for the different pathology samples, comparing bicubic ×2.5 upsampling of the smartphone microscope images and the deep neural network output images.

	Test set	Bicubic upsampling SSIM	Deep neural network SSIM
Masson's-trichrome-stained lung tissue (TIFF input)	90 images (800×800 pixels)	0.4956	0.7020
Masson's-trichrome-stained lung tissue (JPEG input)	90 images (800×800 pixels)	0.5420	0.6830
H&E-stained Pap smear	64 images (1024×1024 pixels)	0.4601	0.7775
Blood smear	9 images (1024×1024 pixels)	0.1985	0.8970

Using the same Masson's-trichrome-stained lung tissue data, the ability of the same neural network to enhance smartphone microscope images that were further degraded by applying lossy compression to them was also evaluated. One important advantage of applying lossy (e.g., JPEG) compression to smartphone microscope images is that compression makes them ideal for storage and transmission/sharing via the bandwidth restrictions of resource-limited environments; this also means that the deep network can perform image enhancement on demand at e.g., the office of a remote pathologist or medical expert. For the smartphone microscope images of the lung tissue, applying JPEG compression reduced an average image with a ~0.1 mm^2 FOV from 1.846 MB to 0.086 MB, resulting in image files that are >21 times smaller. However, lossy compression creates artefacts, such as blocking, and increases the noise and color distortions. As demonstrated in **Figure 1.13**, following the training of the deep network with JPEG-compressed images (**Table 1.2**), it inferred images comparable in quality to those inferred by the deep network that was trained

with lossless compression (TIFF) images. The difference was also assessed using the SSIM and the CIE-94 color distance metrics. As summarized in

Table 1.3 and **Table 1.4**, the average CIE-94 color distance was reduced by approximately 0.067 for the aberration corrected images, while the average SSIM was reduced by approximately 0.02, which form a negligible compromise when scenarios with strict transmission bandwidth and storage limits are considered.

The deep network approach was applied to images of Pap smear samples acquired with the mobile-phone microscope (see **Table 1.2** for implementation details). A Pap smear test is an efficient means of cervical cancer screening, and the sample slide preparation, including its staining, can be performed in a field setting, where a mobile microscope can be of great importance. Due to the thickness of the Pap smear cells (~10–15 μm), imaging such a sample using a high-NA objective lens with a shallow DOF often requires focusing on multiple sample planes. In the training procedure, images from a *single* plane that were acquired by automatic focusing of the benchtop microscope were used. As demonstrated in **Figure 1.14**, the deep network, using the smartphone microscope input images, created enhanced, color-corrected, denoised images with an extended DOF, compared to the images that were captured using the higher-NA objective lens of the benchtop microscope, also emphasized by the yellow arrows in **Figure 1.14**. However, the inexpensive sample holder of the smartphone microscope and its relatively limited axial positioning repeatability makes it challenging to quantify the level of this improvement.

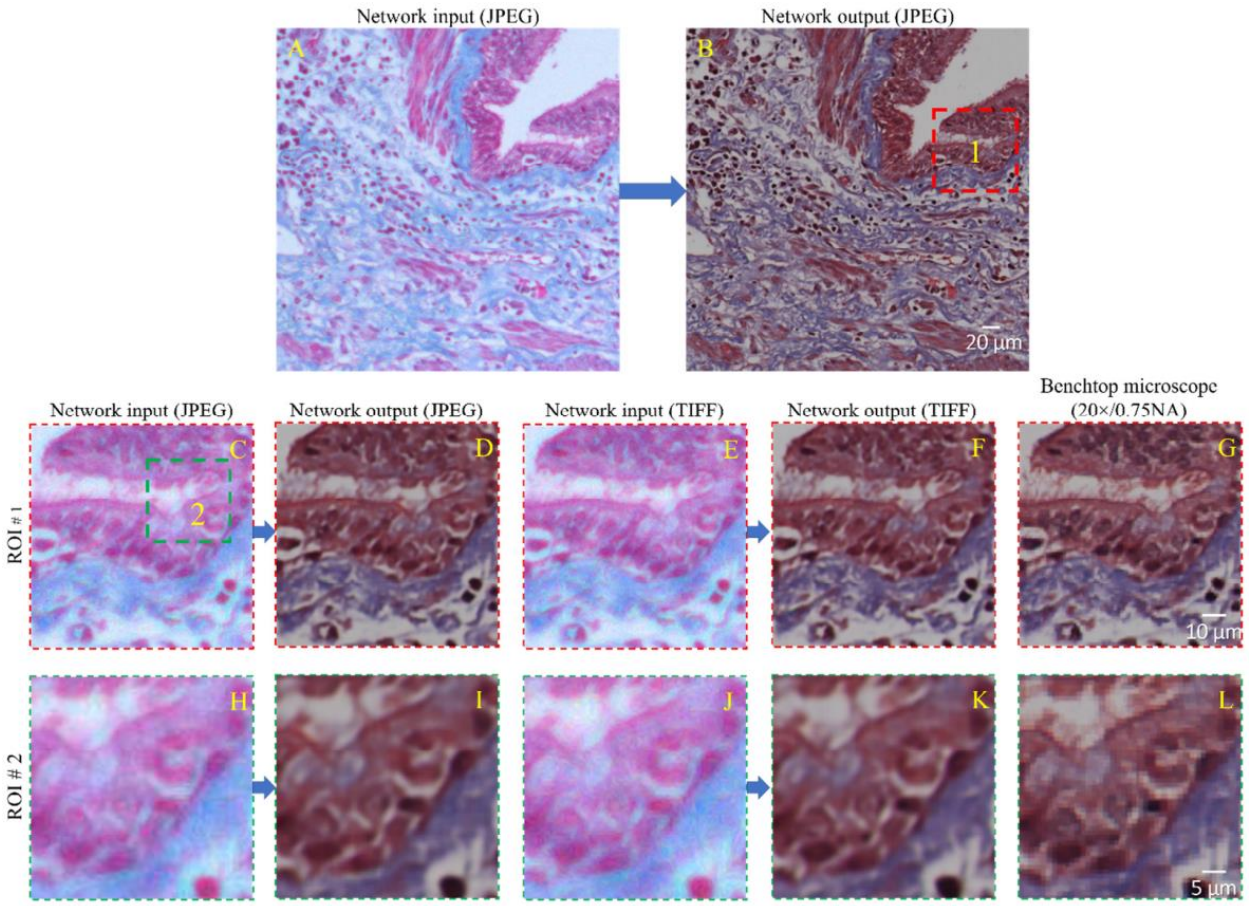

Figure 1.13 Comparison of the deep network inference performance when trained with lossy compression (JPEG) and lossless compression (TIFF). (**A**) JPEG-compressed image, and (**B**) its corresponding deep network output. Zoomed-in versions of (**C**–**G**) ROI #1 and (**H**–**L**) ROI #2.

Some very similar inference results for a human blood smear sample were also obtain as shown in **Figure 1.15**, where the deep network, in a response to an input image of the smartphone microscope (with an average SSIM of ~0.2 and an average color distance of ~20.6) outputs a significantly enhanced image, achieving an average SSIM and color distance of ~0.9 and ~1.8, respectively (see

Table 1.3 and **Table 1.4**).

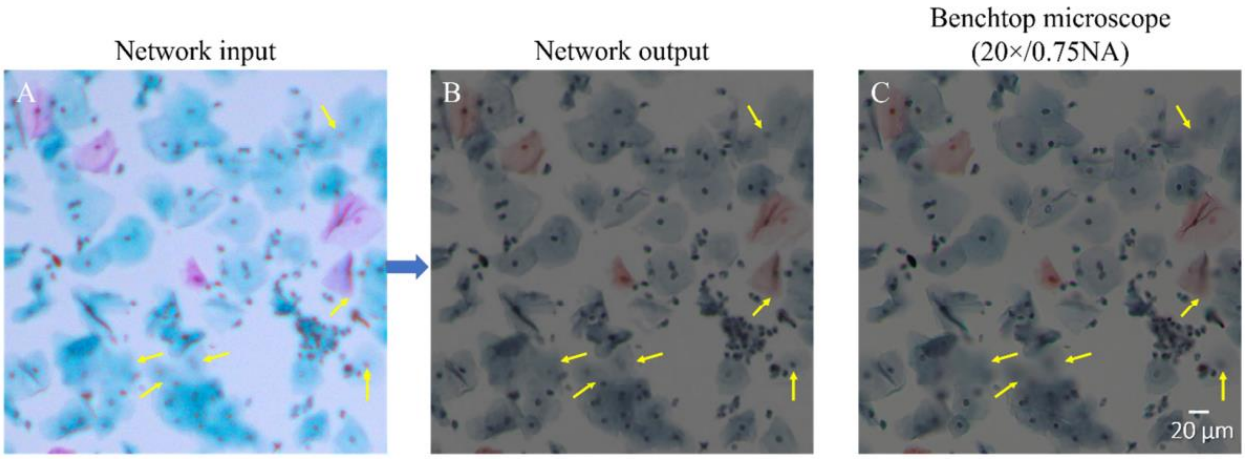

Figure 1.14 Deep neural network output image corresponding to a stained Pap smear sample. (**A**) Smartphone microscope image, (**B**) its corresponding deep network output, and (**C**) a 20×/0.75NA benchtop microscope image. The yellow arrows reveal the extended DOF of the imaging results obtained by the smartphone-based microscope.

While the deep networks were trained with sample-specific datasets in this study, it is possible to train a universal network, at the expense of increasing the complexity of the deep network (for example, increasing the number of channels), which will accordingly increase the inference time and memory resources used [12]. This, however, is not expected to create a bottleneck since image upsampling occurs only in the last two layers in the deep network architecture. Stated differently, the upsampling process is optimized through supervised learning in this approach. Quite importantly, this design choice enables the network operations to be performed in the low-resolution image space, which reduces the time and memory requirements compared with those designs in which interpolated images are used as inputs (to match the size of the outputs) [87]. This design significantly decreases both the training and testing times and relaxes the computational resource requirements, which is important for implementation in resource-limited settings and could pave the way for future implementations running on smartphones. I should also emphasize that the training of multiple mobile-phone microscopes based on the same optical

design can be significantly simplified by using transfer learning [93]. Once a few systems have been trained with the proposed approach, the trained model can be used to initialize the deep network for a new mobile microscope with the already learnt model; this transfer learning-based approach will rapidly converge, even with a relatively small number of example images.

In this work, the smartphone microscope images were captured using the automatic image-capture settings of the phone, which inevitably led the color response of the sensor to be non-uniform among the acquired images. Training the deep network with such a diverse set of images creates a more robust network that will not over-fit when specific kinds of illumination and color responses are present. In other words, the networks that produced generalized, color-corrected responses, regardless of the specific color response acquired by using the automatic settings of the smartphone and the state of the battery-powered illumination component of the mobile microscope. This property should be very useful in actual field settings, as it will make the imaging process more user-friendly and mitigate illumination and image acquisition related variations that could become prominent when reduced energy is stored in the batteries of the illumination module.

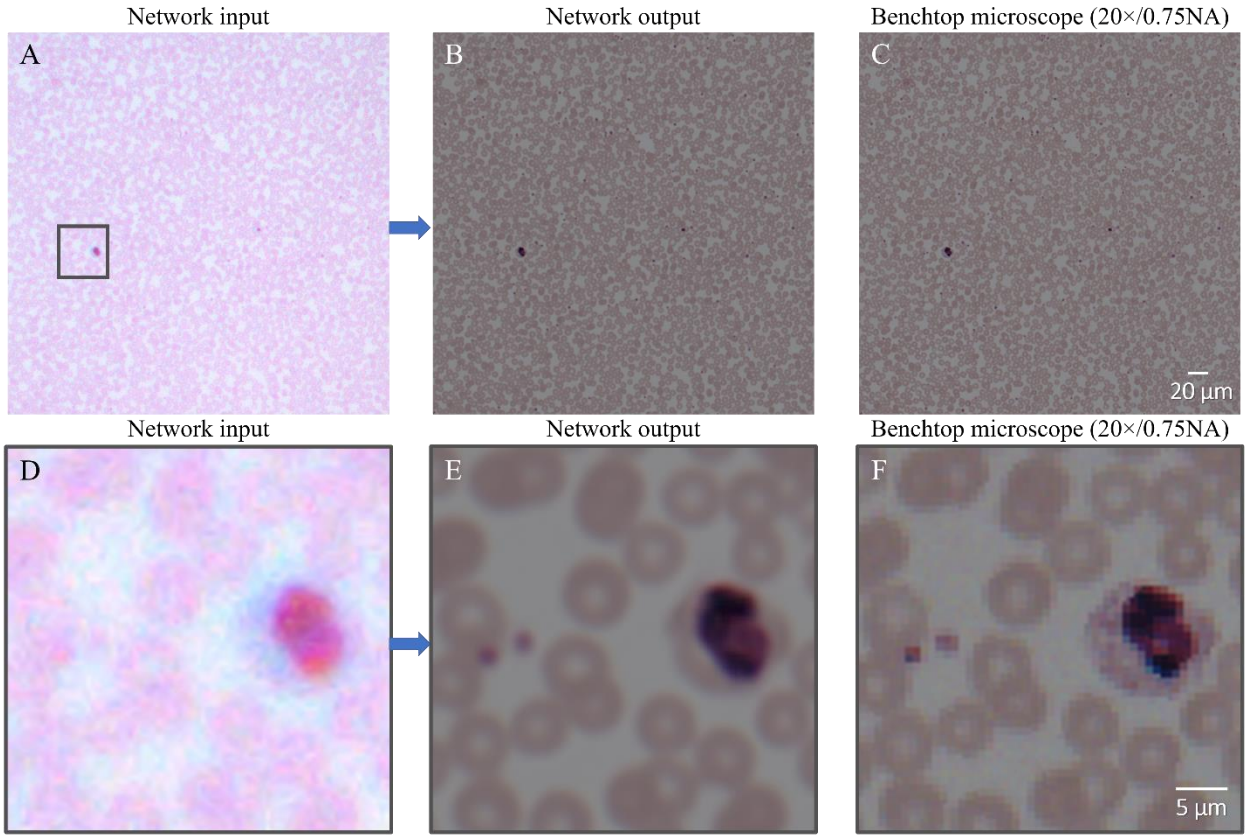

Figure 1.15 Deep neural network output image for a human blood smear sample. (**A**) Smartphone microscope image, (**B**) its corresponding deep network output, and (**C**) a 20×/0.75NA benchtop microscope image of the same sample. (**D**) Zoomed-in version of a ROI of the smartphone microscope image, (**E**) corresponding network output, and (**F**) 20×/0.75NA benchtop microscope image of the same ROI, revealing the image enhancement achieved by the deep neural network.

Furthermore, in recent years, the vast use of digital pathology has highlighted the differences of whole slide pathology images obtained at different laboratories due to the variability in sample preparation, staining procedures, and microscopic image scanning [94]. These variances in color accuracy, resolution, contrast, and dynamic range of the acquired images affect the "fitness for purpose" for diagnostic use, by human observers or automated image analysis algorithms [94]. These issues have created an urgent need for optical image standardization, to better take into account such variations in different stages of the sample preparation, staining as well as

imaging [94]. I believe that the presented deep learning-based approach, with further training, can also be used as part of such an image standardization protocol, by transforming different microscopic images to have similar statistical properties even though they are generated at different laboratories with varying imaging platforms and staining procedures. This would help standardize the images obtained by various cost-effective and mobile microscopes, further enhance their spread and use in biomedical and clinical applications and reduce diagnostic discrepancies that might result due to above discussed variations in the raw acquired images. Once an image standard has been decided by a group of experts, calibration slides and procedures can be created for acquiring images using different microscopy systems being used in different settings, and all these images can be used to train local and universal deep neural networks that can enhance a given input image to the desired standard.

Although smartphone microscopes possess certain advantages, such as integration with off-the-shelf consumer products benefiting from economies of scale, portability, and inherent data communication, a plethora of other devices and platforms (e.g., Raspberry Pi) with different capabilities can be employed as cost-effective microscopes and benefit from the presented deep learning-based approach. For example, by using a compact benchtop microscope composed of cost-effective objective lenses and illumination sources, some of the mechanical (e.g., related to object holder and its alignment) and illumination instabilities should produce less degradation in image quality than that resulting from using a smartphone-based mobile microscope. Such an imaging apparatus with its better repeatability in imaging samples will facilitate the use of the pyramid elastic registration as part of the image enhancement workflow, since the image distortions will be more stationary and less affected by mechanical and illumination instabilities resulting from, e.g., user variability and the status of the battery. For that, one could use the average

block-shift correction maps calculated between the high-end and cost-effective microscope images; for example, see the mean shift map calculated for the FOV of the lung tissue sample.

Conclusion

This research demonstrates the proof-of-concept of a deep learning-based framework to enhance mobile-phone microscopy by creating high-resolution, denoised and color-corrected images through a convolutional neural network (CNN). Clinical validation is left outside the scope of this manuscript; however, in future, I plan to test the presented approach and the resulting image enhancement through a randomized clinical study to validate and quantify its impact on medical diagnostics and telemedicine related applications.

To conclude, the significant enhancement of low-resolution, noisy, distorted images of various specimens acquired by a cost-effective, smartphone-based microscope by using a deep learning approach was demonstrated. This enhancement was achieved by training a deep convolutional neural network using the smartphone microscope images and corresponding benchtop microscope images of various specimens, used as gold standard. The results, which were obtained using a non-iterative feed-forward (i.e., non-cyclic) algorithm, exhibited important advantages such as the enhancement and restoration of fine spatial features, correction for the colour aberrations, and removal of noise artefacts and warping, introduced by the mobile phone microscope optical hardware/components. For samples that naturally include height/depth variations, such as Pap smear samples, the advantage of DOF extension with respect to the images of a benchtop microscope with a higher NA was also observed. These results demonstrate the potential of using smartphone-based microscopes along with deep learning to obtain high-quality images for telepathology applications, relaxing the need for bulky and expensive microscopy equipment in resource-limited settings. Finally, this presented approach might also provide the

basis for a much-needed framework for standardization of optical images for clinical and biomedical applications.

Chapter 2 Deep learning enables cross-modality super-resolution in fluorescence microscopy

2.1 Introduction

Deep learning-enabled image super-resolution of smartphone microscope images presented in Chapter 1 demonstrates the capability of network-based cross-modality image transformation: from a smartphone imaging platform to a high-end benchtop imaging platform. I further explored the cross-modality possibilities in fluorescence microscopy and demonstrate even more exciting progress in this chapter. Super-resolution microscopy methods such as localization microscopy [95–98], stimulated emission depletion (STED) microscopy [99], and structured illumination microscopy (SIM) [100–102] provide unprecedented access to the inner workings of cells and various biological processes. However, these methods often rely on relatively sophisticated optical setups, specific fluorophores and mounting media, and extensive computational post-processing of acquired image data [103–105], which in and of itself may require *a priori* knowledge about the sample and/or its preparation as well as a physical model of the image formation process [106–109], including, for example, the point-spread-function (PSF) of the imaging system. In general, more accurate models yield higher quality results, often with a trade-off of exhaustive parameter search and computational cost.

Here I will introduce a deep learning-based framework to achieve super-resolution and cross-modality image transformations in fluorescence microscopy without the need for making any assumptions on or modeling of the image formation process. I train a deep neural network using a Generative Adversarial Network (GAN) [57] model to transform an acquired low-resolution image into a high-resolution one using matched pairs of experimentally acquired low and higher resolution images. The success of this super-resolution approach is a result of a highly-accurate

multi-stage image registration and alignment process (discussed in the 2.6 Materials and methods) between the lower resolution and the corresponding higher resolution images, which allows the network to solely focus on the task of improving the resolution of a previously unseen input image.

Once the deep network is trained, it remains fixed and can be used to rapidly output batches of high-resolution images, in e.g., 0.4 sec for an image size of 1024×1024 pixels using a single Graphics Processing Unit (GPU). The network inference is non-iterative and does not require a manual parameter search to optimize its performance.

I will demonstrate the success of this deep learning-based framework by improving the resolution of raw images captured by different imaging modalities, including wide-field fluorescence, confocal, and TIRF microscopes. In the wide-field imaging case, the images acquired using a 10×/0.4NA objective lens are transformed into resolution-enhanced images that match the images of the same samples acquired with a 20×/0.75NA objective. In the second case, I perform cross-modality transformation of diffraction-limited confocal microscopy [110] images to match the images that were acquired using a STED microscope [95,99], super-resolving Histone 3 distributions within HeLa cell nuclei and also showing a PSF width that is improved from ~290 nm down to ~110 nm. As another example of this GAN-based cross-modality image transformation framework, time-lapse TIRF microscopy images were super-resolve to match TIRF-SIM [111] images of endocytic clathrin-coated structures in SUM159 cells and Drosophila embryos. This deep learning-based fluorescence super-resolution approach improves both the field-of-view (FOV) and imaging throughput of fluorescence microscopy and can be used to transform lower-resolution and wide-field images acquired using various imaging modalities into higher resolution ones.

Part of this chapter has been previously published in :

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

2.2 Resolution enhancement in wide-field fluorescence microscopy

I initially demonstrated the resolution improvement of the presented approach by imaging bovine pulmonary artery endothelial cell (BPAEC) structures. In the training stage, for each excitation line (DAPI, FITC, and TxRed) I used a multi-stage image registration process to accurately align 2625 pairs of low- and high-resolution image patches to each other, and a separate model was trained for each filter set to achieve optimal results (see 2.6 Materials and methods). Each image patch had a size of 1024×1024 pixels, and the raw input images to the network were acquired using a 10×/0.4NA objective and the results of the network were compared against the ground truth images, which were captured using a 20×/0.75NA objective. An example of the network input image is shown in **Figure 2.1**a, where the FOV of the 10× and 20× objectives are also labeled. **Figure 2.1**b,c show some zoomed-in regions-of-interest (ROIs) revealing further details of a cell's F-actin and microtubules. A pretrained deep neural network is applied to each color channel of these input images (10×/0.4NA), outputting the resolution-enhanced images shown in **Figure 2.1**d,e, where various features of F-actin, microtubules, and nuclei are clearly resolved at the network output, providing a very good agreement to the ground truth images (20×/0.75NA) shown in **Figure 2.1**f,g. Note that all the network output images shown in this manuscript were blindly generated by the deep network, i.e., the input images were not previously seen by the network.

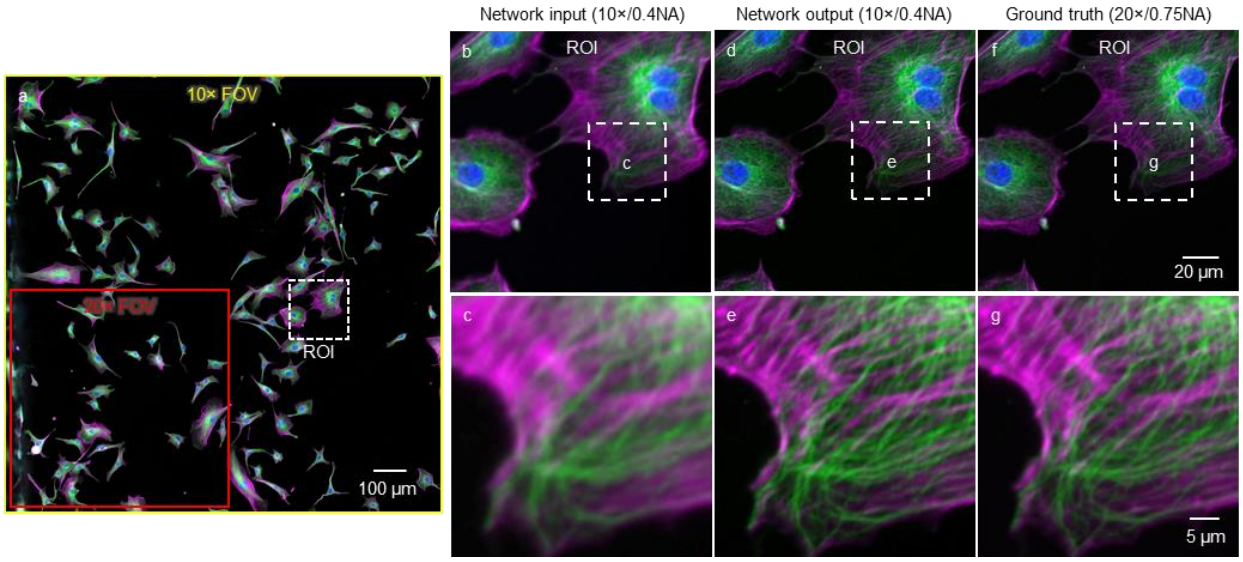

Figure 2.1 Deep-learning-based super-resolved images of bovine pulmonary artery endothelial cells (BPAECs). **a**, Network input image acquired with a 10×/0.4-NA objective lens. **b**–**g**, Smaller ROIs are magnified and shown in (**b**,**c**) network input, (**d**,**e**) network output, and (**f**,**g**) ground truth (20×/0.75-NA). Experiments were repeated with >250 images, achieving similar results. Color map: magenta for F-actin, green for microtubules, blue for nuclei.

Next, I compared the results of deep learning-based super-resolution against widely-used image deconvolution methods, i.e., the Lucy-Richardson (LR) deconvolution and the non-negative least square (NNLS) algorithm. [112–114] For this, I used an estimated model of the PSF of the imaging system, which is required by these deconvolution algorithms to approximate the forward model. Following its parameter optimization (2.6 Materials and methods), the LR deconvolution algorithm, as expected, demonstrated resolution improvements compared to the input images (**Figure 2.2**a,f,k); however compared to the deep learning results (**Figure 2.2**b,g,l), the improvements observed with LR deconvolution (**Figure 2.2**c,h,m) are modest, despite the fact that it used parameter search, optimization and *a priori* knowledge on the PSF of the imaging system. The NNLS algorithm, on the other hand, yields slightly sharper features (see **Figure 2.2**d,i,n) compared to LR deconvolution results, at the cost of having additional artifacts; regardless, both

of these deconvolution methods are inferior to the deep learning results reported in **Figure 2.2**, exhibiting a shallower modulation depth in comparison to the deep learning results and the ground truth images.

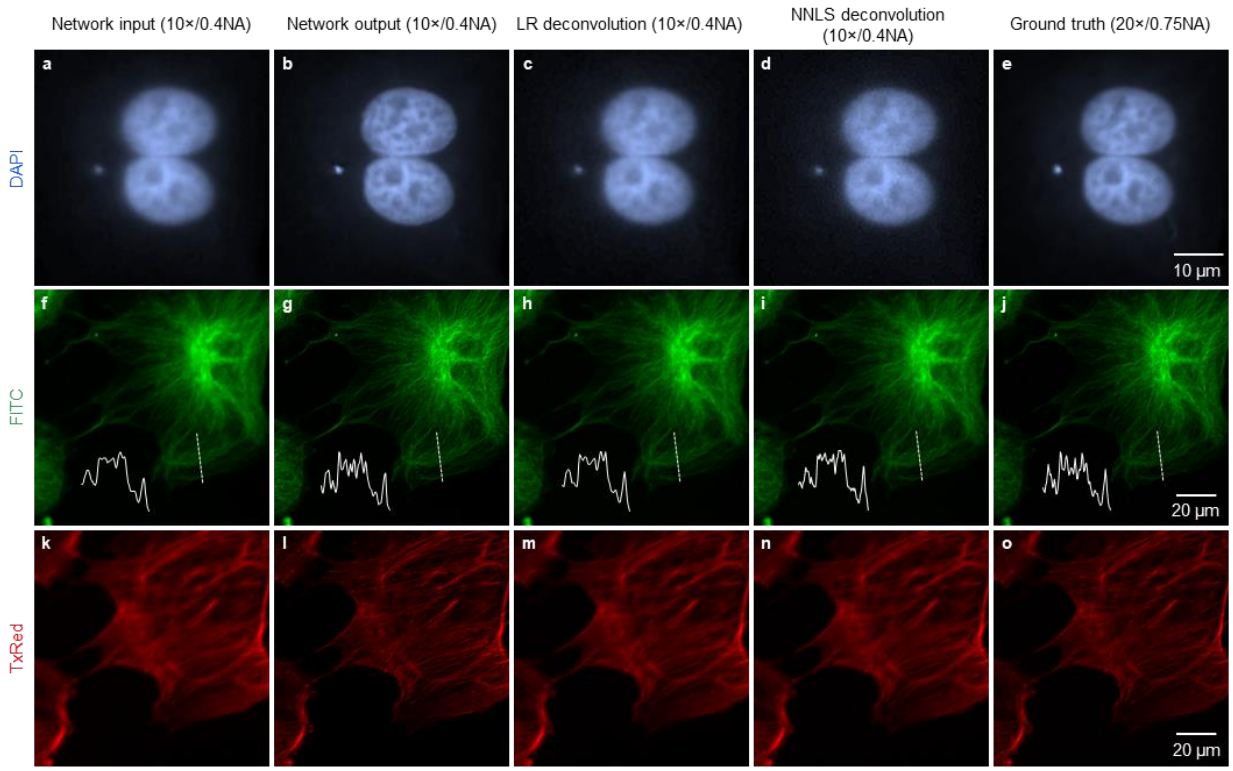

Figure 2.2 Comparison of deep learning results against Lucy–Richardson (LR) and non-negative least square (NNLs) image deconvolution algorithms.

I also noticed that the deep network output image shows sharper details compared to the ground truth image, especially for the F-actin structures. This result is in-line with the fact that all the images were captured by finding the autofocusing plane within the sample using the FITC channel (see e.g., **Figure 2.2**f-j), and therefore the Texas-Red channel (e.g., **Figure 2.2**k-o) can remain slightly out-of-focus due to the thickness of the cells. This means the shallow depth-of-field (DOF) of a 20×/0.75NA objective (~1.4 μm) might have caused some blurring in the F-actin structures (**Figure 2.2**o). This out-of-focus imaging of different color channels is not impacting

the network output as much since the input image to the network was captured with a much larger DOF (~5.1 μm), using a 10×/0.4NA objective. Therefore, in addition to an increased FOV resulting from a low NA input image, the network output image is also benefiting from an increased DOF, helping to reveal some finer features that might be out-of-focus in different color channels using a high NA objective.

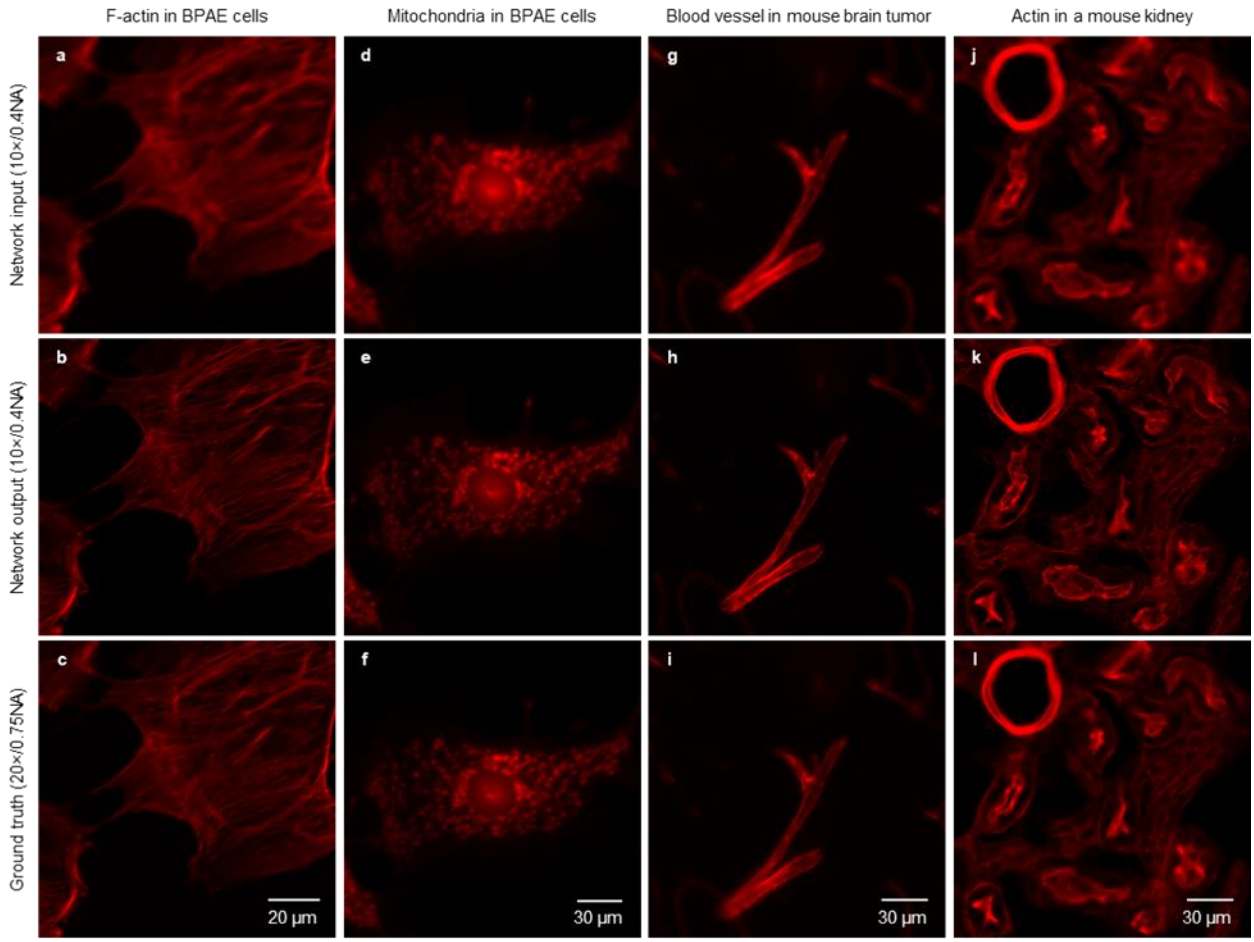

Figure 2.3 Generalization of a neural network model trained with F-actin to new types of structures that it was not trained for. Network input, output, and ground truth images corresponding to (a-c) F-actin inside a BPAEC (image not in the training dataset), (d-f) mitochondria inside a BPAEC, (g-i) blood vessel in mouse brain tumor, and (j-l) actin in a mouse kidney section demonstrate that all these structures can be blindly super-resolved by a neural network that was trained with only F-actin images. Experiments were repeated with >50 images with similar results.

Next, I tested the generalization of the trained network model in improving image resolution on new types of samples that were not present in the training phase; **Figure 2.3** demonstrates the resolution enhancement when applying the network model trained with F-actin (**Figure 2.3**a-c) to super-resolve images of mitochondria in BPAEC (**Figure 2.3**d-f), blood vessels in a mouse brain tumor (**Figure 2.3**g-i), and actin in a mouse kidney tissue (**Figure 2.3**j-l). Even though these new types of objects were not part of the network's training set, the deep network was able to correctly infer their fine structures through blind inference. In these experiments both the training and the blind testing images were taken with the same fluorescence filter set. **Figure 2.3**e,f,h,i,k,l, further support the enhanced DOF of the network output images for various types of samples when compared to the ground truth, higher NA images. Once again, I want to emphasize that a new network model should be trained for achieving optimal super-resolution performance on input images corresponding to different types of samples or captured with a new experimental setup. However, in case such training image pairs are not available to follow the super-resolution image transformation framework, one can attempt to use an existing trained model, although this might not produce ideal results in all cases.

For wide-field microscopy images, I performed quantification of the deep network results using spatial frequency spectrum analysis: in **Figure 2.4** I compared the spatial frequency spectrum of the network output images (for BPAEC structures) with respect to the network input images to demonstrate the frequency extrapolation nature of the deep learning framework. The cross-section of the radially-averaged power spectrum confirms the success of the network output, matching the extended spatial frequency spectrum that is expected from a higher-resolution imaging system (as illustrated with the overlap of the red and orange curves in **Figure 2.4**g).

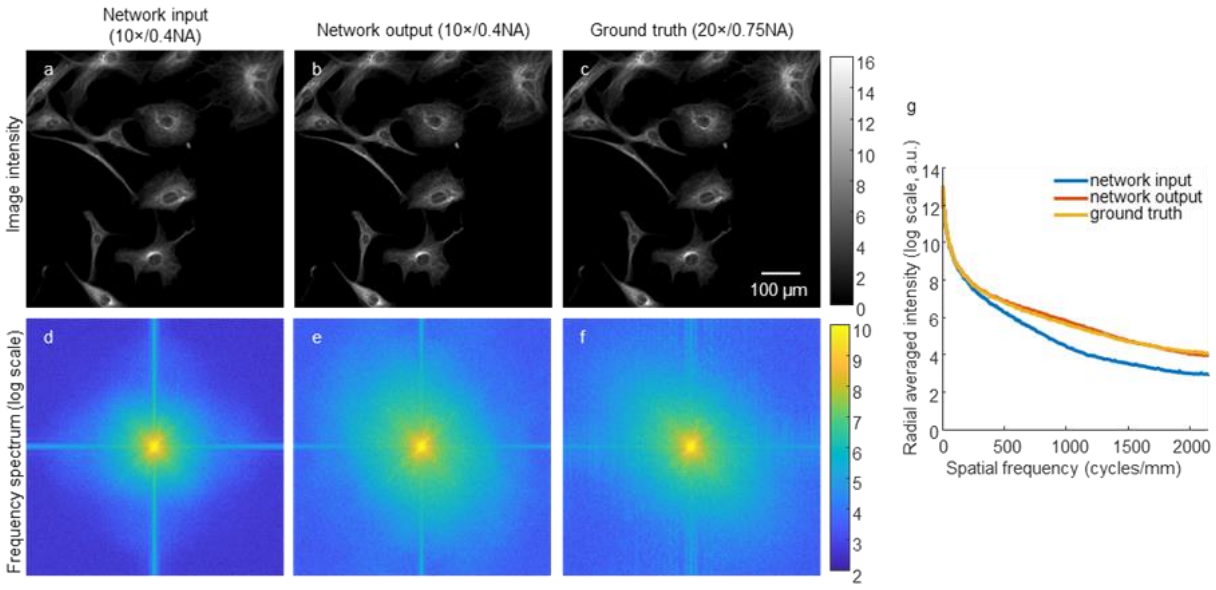

Figure 2.4 Illustration of the spatial frequency extrapolation achieved by deep learning. The deep learning model takes (a) an input image of microtubules in BAPEC obtained using a 10×/0.4NA objective lens and super-resolves it as shown in (b), to match the resolution of (c) the ground truth image which is acquired with a 20x/0.75NA objective lens. (d-f) show the spatial frequency spectra in log scale, corresponding to (a-c), respectively. (g) shows the radially averaged intensity of each one of the spatial frequency spectra shown in (d,e,f). Analysis was performed on a randomly selected image from a group of 94 images with similar results.

I further quantified the resolution improvement achieved in wide-field images using the deep learning approach by imaging 20 nm fluorescent beads at an emission wavelength of 645 nm (see 2.6 Materials and methods) and used the images acquired with a 10×/0.4NA objective lens as input to the deep network model, which was trained *only* with F-actin (as demonstrated in **Figure 2.1** and **Figure 2.2**). The super-resolution results of the deep network are summarized in **Figure 2.5**. To quantify the resolution improvement in these results, I measured the PSFs arising from the images of single/isolated nano-beads across the imaging FOV [115]; this was repeated for >100 individual particles that were tracked in the network input and output images, as well as the ground truth images (acquired using a 20×/0.75NA objective lens). The full-width at half-maximum

(FWHM) of the 10× input image PSF is centered at ~1.25μm, corresponding to a sampling rate limited by an effective pixel size of ~0.65 μm. Despite the fact that the fluorescent signal from 20 nm beads is rather weak, the deep neural network (trained only with BPAEC samples) successfully picked up the signal from individual nano-beads and blindly improved the resolution to match that of the ground truth, as shown in the PSF comparison reported in **Figure 2.5**d. These results further highlight the robustness of the deep learning method to low SNR (signal-to-noise ratio) as well as its generalizability to different spatial structures of the object. The broadening of the PSF distribution in 20×/0.75NA images (see **Figure 2.5**d) can be attributed to the smaller DOF of the high NA objective lens, where the nano-beads at slightly different depths are not in perfect focus and therefore result in varying PSF widths. The deep network results, on the other hand, once again demonstrate the enhanced DOF of the network output image, showing uniform focusing with improved resolution at the network output image.

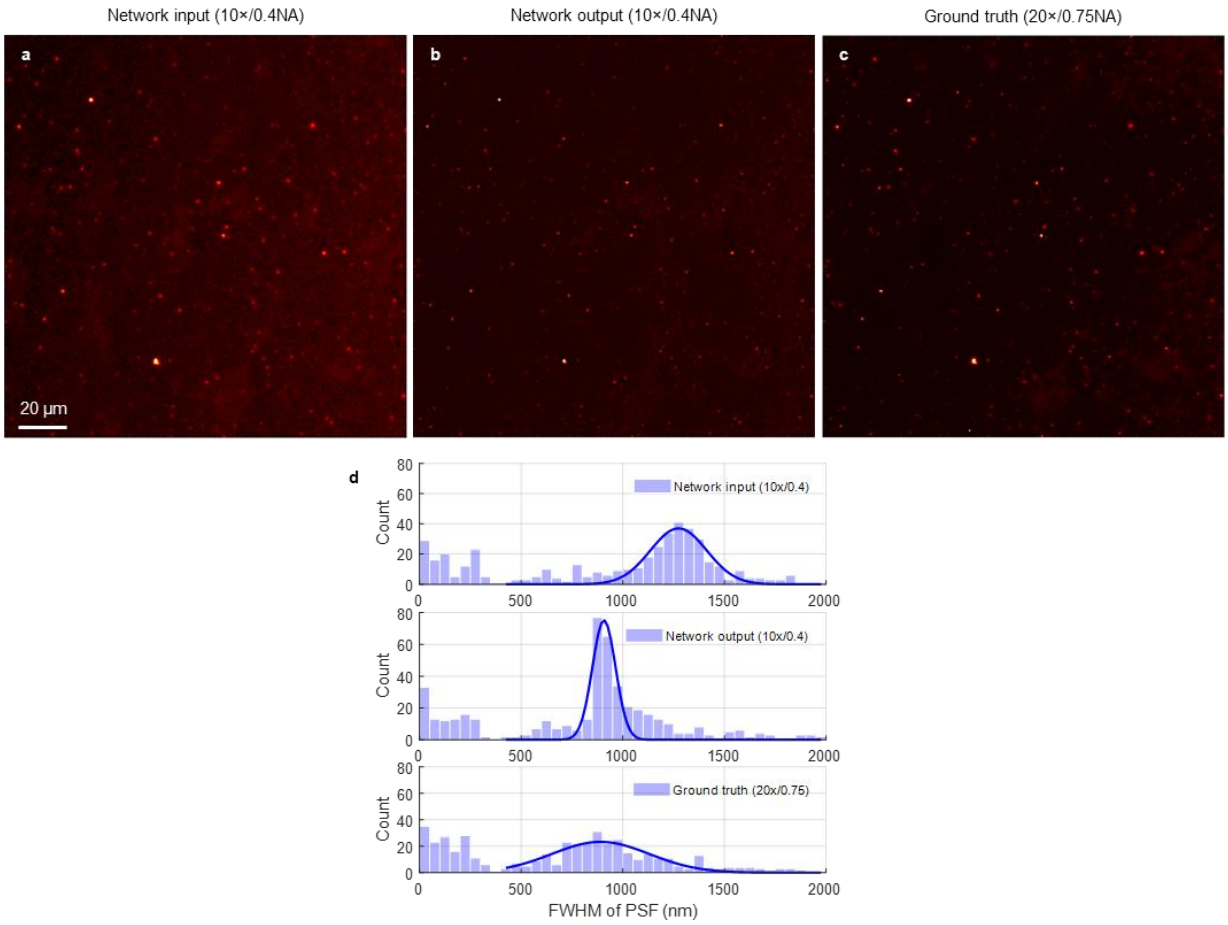

Figure 2.5 PSF characterization of wide-field images.(a) An example image of 20 nm fluorescent particles captured with a 10×/0.4NA objective lens as the neural network input. (b) The network inference image with a model pre-trained with only F-actin images. (c) The ground truth image captured with a 20×/0.75NA objective lens. (d) PSF characterization, before and after the network inference, and its comparison to the ground truth image. I extracted more than 200 bright spots from the same locations of the network input (10×/0.4NA), network output (10×/0.4NA), and the corresponding ground truth (20×/0.75NA) images. Each one of these spots was fit to a 2D Gaussian function and the corresponding FWHM distributions are shown in each histogram. Analysis was performed over 3 different images randomly selected from the same nanobead sample.

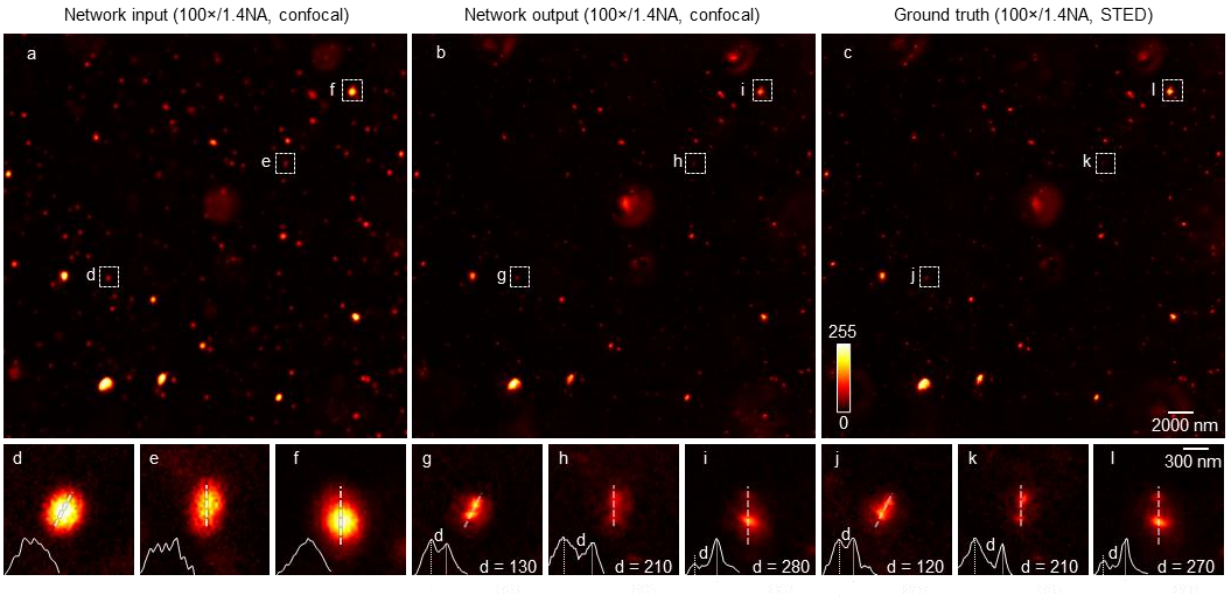

Figure 2.6 Image resolution improvement beyond the diffraction limit: from confocal microscopy to STED. a–c, A diffraction-limited confocal microscope image is used as input to the network and is super-resolved to blindly yield (b) the network output, which is comparable to (c) a STED image of the same FOV, used as the ground truth. d–f, Examples of closely spaced nano-beads that cannot be resolved by confocal microscopy. g–l, the trained neural network takes d–f as input and resolves the individual beads (g–i), very well agreeing with STED microscopy images (j–l). The cross-sectional profiles reported in d–l are extracted from the original images. Peak-to-peak distance (d) in these cross-sectional profiles is reported in nanometers. Also see **Figure 2.7** for further quantification of the performance of the deep network on confocal images, and its comparison to STED. Experiments were repeated with 75 images, achieving similar results.

2.3 Cross-modality imaging from confocal to STED

I also applied the presented framework to transform confocal microscopy images into images that match those obtained by STED microscopy (**Figure 2.6**, **Figure 2.7**). Training data were acquired using 20 nm fluorescent beads (645 nm emission) imaged on the same instrument using both confocal microscopy and STED modes. After the training phase, the neural network, as before, blindly takes an input image (confocal) and outputs a super-resolved image that matches

the STED image of the same sample. Some of the nano-beads in my samples were spaced close to each other, within the classical diffraction limit, i.e., under ~290 nm, as shown in e.g., **Figure 2.6**d-f, and therefore could not be resolved in the raw confocal microscopy images. The neural network resolved these closely-spaced nano-particles, providing a good match to STED images of the same regions of the sample, see **Figure 2.6**g,h,i vs. **Figure 2.6**j,k,l.

To further quantify this resolution improvement achieved by the network, I measured the PSFs arising from the images of single/isolated nano-beads across the sample FOV [115] following the same method described earlier, repeated for >400 individual nanoparticles that were tracked in the images of the confocal microscope and STED microscope, as well as the network output image (in response to the confocal image). The results are summarized in **Figure 2.7**, where the FWHM of the confocal microscope PSF is centered at ~290 nm, roughly corresponding to the lateral resolution of a diffraction-limited imaging system at an emission wavelength of 645 nm. As shown in **Figure 2.7**, PSF FWHM distribution of the network output provides a very good match to the PSF results of the STED system, with a mean FWHM of ~110 nm vs. ~120 nm, respectively.

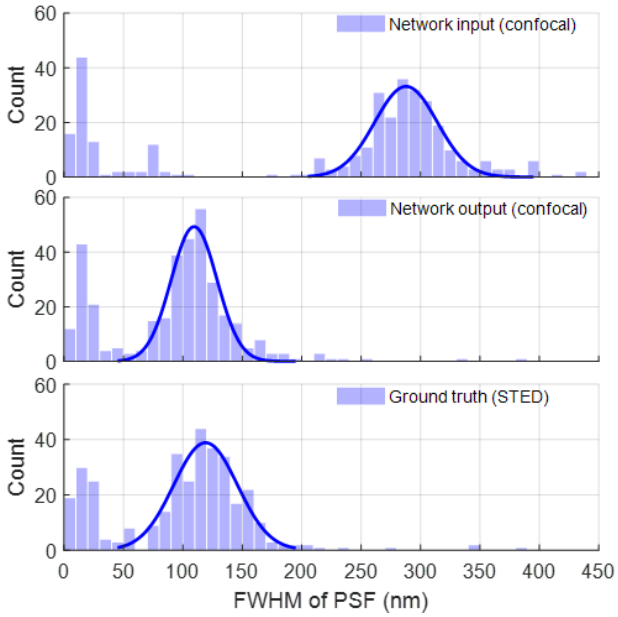

Figure 2.7 PSF characterization, before and after the network, and its comparison to STED. More than 400 bright spots from the same locations of the network input (confocal), network output (confocal), and the corresponding ground truth (STED) images were extracted. Each one of these spots was fit to a 2D Gaussian function, and the corresponding FWHM distributions are shown in each histogram. These results show that the resolution of the network output images is significantly improved from ~290 nm (top row: network input using a confocal microscope) to ~110 nm (middle row: network output), which provides a very good fit to the ground truth STED images of the same nano-particles, summarized in the bottom row.

An additional benefit of using the deep learning approach is improved SNR, for which I conducted a comparative analysis using the confocal-to-STED transformation results to quantify this improvement. For this analysis, I selected a small FOV containing a single 20 nm bead and calculated the SNR for the network input (confocal image), the network output and the ground truth image (STED). The SNR is defined as:

$$SNR = \left| \frac{s - \bar{b}}{\sigma_b} \right| \tag{2.1}$$

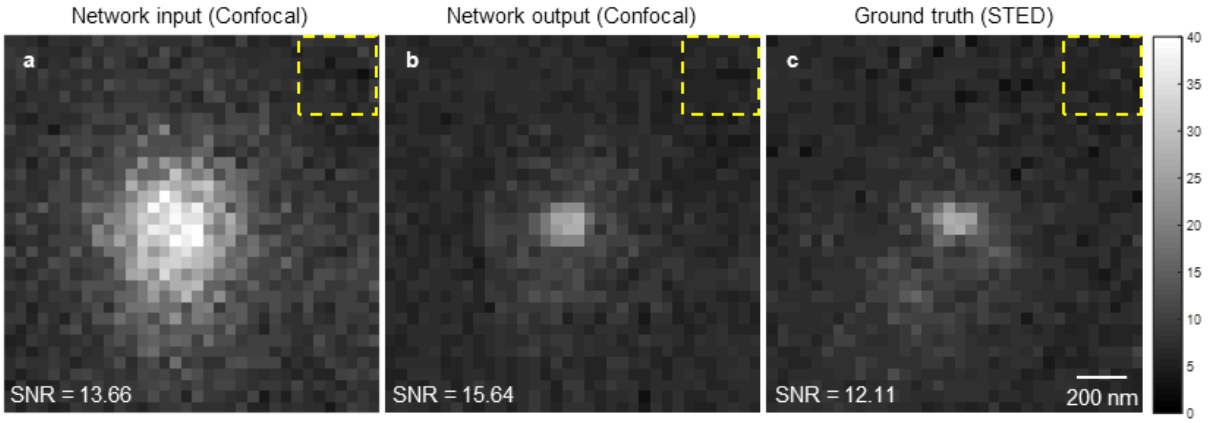

Figure 2.8 Quantification of the SNR improvement achieved by the confocal-to-STED transformation network. (a) Input image SNR= 13.66. (b) Network output image SNR=15.64. (c) STED image SNR= 12.11. The yellow dashed line regions are used to calculate the background mean and variation. Analysis was performed on a randomly selected particle from a group of 75 images with similar results.

where s is the peak value of the signal calculated from a Gaussian fit to the particle (see 2.6 Materials and methods), $\bar{b}$ is the mean value of the background (e.g. the regions defined with the yellow dashed lines in **Figure 2.8**), σ_b is the standard deviation of the background. The results shown in **Figure 2.8** reveal that the deep neural network suppresses noise and improves the SNR compared to the input image as well as the ground truth image (STED).

Next, I applied this confocal-to-STED image transformation framework to super-resolve Histone 3 distributions within fixed HeLa cell nuclei (see **Figure 2.9**). Because nanoparticles do not accurately represent the spatial feature diversity observed in biological specimens, direct application of a network that is trained only with nanobeads would not be ideal to image complex biological systems. Therefore, I made use of a concept known as "transfer learning" [116], in which a learned neural network (trained e.g., with nanoparticles, **Figure 2.6**, **Figure 2.7**) was used to initialize a model to super-resolve cell nuclei using confocal-to-STED transformation; this

transfer learning approach also significantly speeds up the training process as detailed in Online **Methods** section. Despite some challenges associated with STED imaging of densely labeled specimen as well as sample drift, after transfer learning, the neural network successfully improved the resolution of a confocal microscope image (input), matching the STED image of the same nuclei (**Figure 2.9**). Some of the discrepancies between the network output and the STED image can be related to the fluctuations observed in STED imaging, as shown in **Figure 2.9**d-f, where 3 consecutive STED scans of the same FOV show frame-to-frame variations due to fluorophore state changes and sample drift. In this case, the network's output image better correlates with the average of three STED images that are drift-corrected (see **Figure 2.9**b,c). Using the same confocal-STED experimental data, **Figure 2.11** further illustrates the advantages of the presented GAN-based super-resolution approach over a standard CNN (convolutional neural network) without the discriminative loss, which results in a lower resolution image compared to GAN-based inference.

I should also emphasize that, in the experiments reported in **Figure 2.6**, **Figure 2.7**, and **Figure 2.9**, the required excitation power for STED was 3-10-fold stronger than that of confocal microscopy (see 2.6 Materials and methods). Furthermore, the depletion beam of STED is typically orders-of-magnitude higher than its excitation beam, [117–119] which highlights an important advantage of the deep learning-based super-resolution approach for imaging biological objects that are vulnerable to photo-bleaching or photo-toxicity. [117,120]

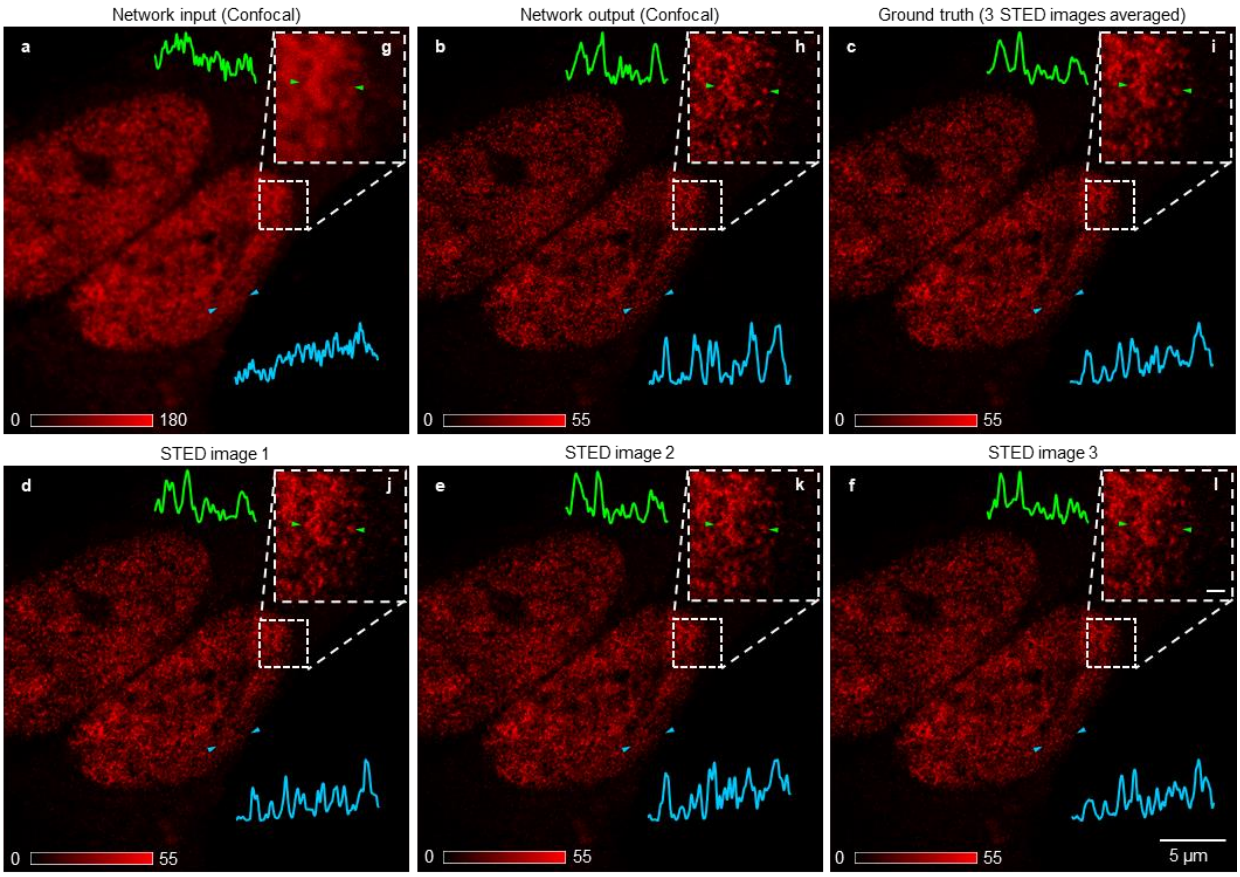

Figure 2.9 Deep-learning enabled cross-modality image transformation from confocal to STED. (a) A diffraction-limited confocal microscope image of Histone 3 distributions within HeLa cell nuclei is used as input to the neural network to blindly yield (b) the network output image, which is comparable to (c) STED image of the same FOV. Figure (c) is the average of 3 individual STED scans of the same FOV, shown in (d,e,f) respectively. Scale bar in (l) is 500 nm. Arrows in each image refer to the line of the shown cross-section. Experiments were repeated with 30 images achieving similar results.

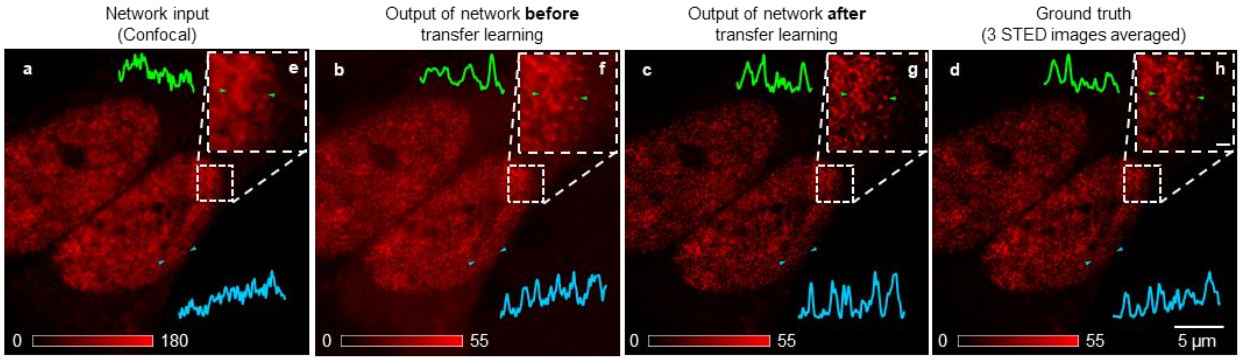

Figure 2.10 A neural network model trained with nano-bead images exhibits significantly improved performance in blindly inferring Histone 3 distributions within fixed HeLa cell nuclei after applying transfer learning with similar images. (a) Network input image captured with a confocal microscope. (b) Network inference image by a model pre-trained only with fluorescent particle images. (c) Network inference image by a model pre-trained only with fluorescent particle images and then transfer learnt with cell nuclei images. (d) The ground truth image captured with a STED microscope. (e-h) Zoomed-in regions (a-d). Scale bar in (h) is 500 nm. Arrows in each image point to the line of the shown cross-section.

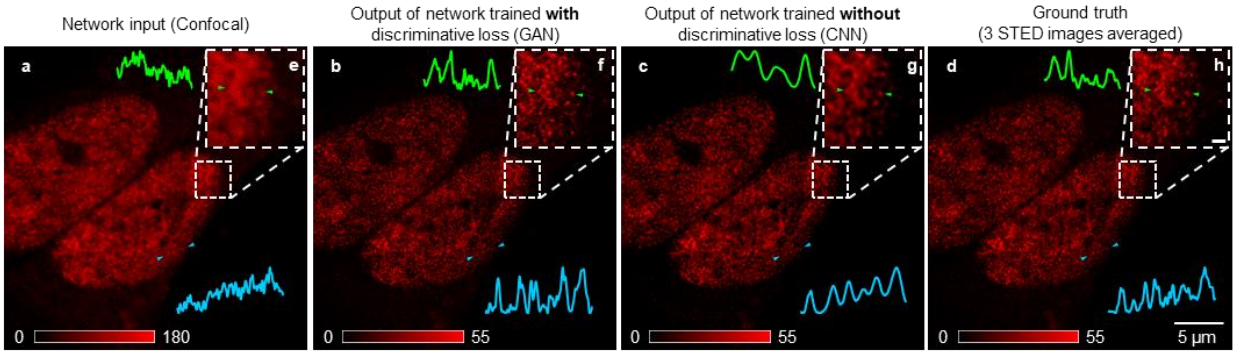

Figure 2.11 Discriminative loss is critical to the training of a generative network. (a) Network input image captured with a confocal microscope. (b) Network inference image by the same generative model as in **Figure 2.9**, trained with the discriminative loss, i.e., the GAN framework. (c) Network inference image by the same generative model as in **Figure 2.9**, trained without the discriminative loss, shows compromised performance compared to (b). (d) The ground truth image captured with a STED microscope. (e-h) Zoomed-in regions (a-d). (c) and (g) show over-smoothed structures and missing details. Scale bar in (h) is 500 nm. Arrows in each image refer to the line of the shown cross-section.

2.4 Cross-modality imaging from TIRF to TIRF-SIM

I further demonstrated the cross-modality image transformation capability of my method by transforming diffraction-limited TIRF images to match TIRF-SIM reconstructions (**Figure 2.12** and **Figure 2.13**). In these experiments, the sample was exposed to 9 different structured illumination patterns following a reconstruction method used in SIM [111], whereas the low-resolution (diffraction-limited) TIRF images were obtained using a simple average of these 9 exposures [121]. I trained a neural network model using images of gene-edited SUM159 cells expressing eGFP-labeled clathrin adaptor AP2, and blindly tested its inference (**Figure 2.12**). To highlight some examples, the neural network was able to detect the dissociation of clathrin-coated pits from larger clathrin patches (i.e. plaques [111,122]) as shown in **Figure 2.12**r,t as well as the development of curvature-bearing clathrin cages [111,123], which appear as doughnuts under SIM (**Figure 2.12**l-o). Next, to provide another demonstration of the network's generalization, it was blindly applied to amnioserosa tissues of Drosophila embryos (never seen by the network) expressing clathrin-mEmerald (**Figure 2.13**). Highly motile clathrin-coated structures [124] within the embryo that cannot be resolved in the original TIRF image can be clearly distinguished as separate objects in the network output (**Figure 2.13**). These results demonstrate that the network model can super-resolve individual clathrin-coated structures within cultured cells and tissues of a developing metazoan embryo.

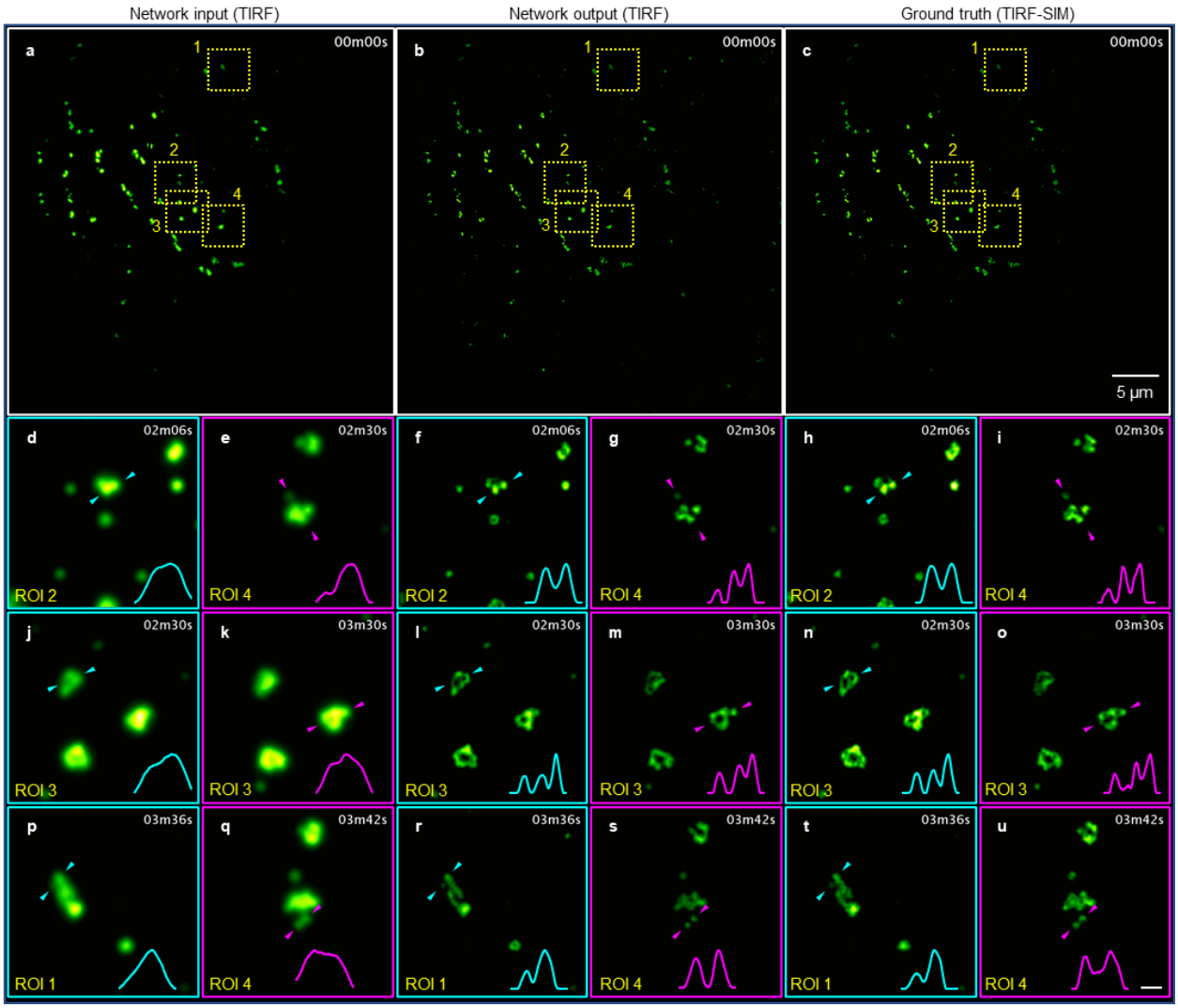

Figure 2.12 Deep-learning enabled cross-modality image transformation from TIRF to TIRF-SIM. (a) TIRF image of a gene edited SUM159 cell expressing AP2-eGFP. (b) The network model super-resolves the diffraction-limited TIRF image (input) and matches (c) TIRF-SIM reconstruction results. (d-u) Zoom-in regions of (a-c) at the labeled ROIs and time points. Scale bar in (u) is 500 nm. Arrows in each image refer to the line of the shown cross-section. Experiments were repeated with >1000 images achieving similar results.

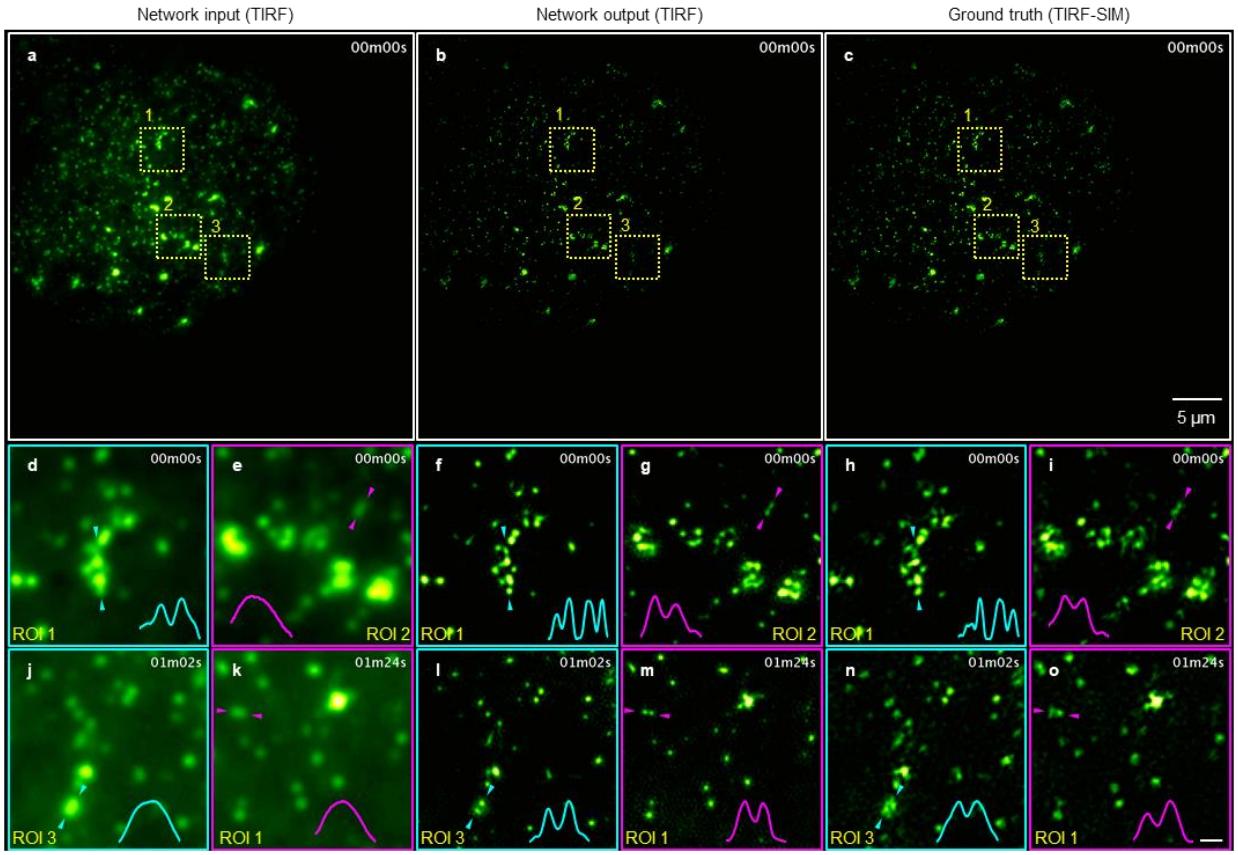

Figure 2.13 Super-resolution imaging of amnioserosa tissues of a Drosophila embryo expressing Clathrin-mEmerald using the TIRF to TIRF-SIM transformation network that was trained only with AP2 images (a) Diffraction-limited TIRF image as network input. (b) Network inference image by a model-pretrained only with AP2 images. (c) Ground truth image by SIM remonstration. (d-o) Comparison of enlarged ROIs at different time points shows super-resolved details of the amnioserosa tissues. The capturing time point is labeled on the upper-right corner of each image. These results provide additional examples of the generalization of the network's inference to new sample types that it has never seen before. To position the apical surface of amnioserosa cells within the evanescent excitation field of the TIRF system, the dechorionated embryo was gently pressed against the coverglass. The relatively high levels of reconstruction artifacts observed in the TIRF-SIM images can be attributed to the autofluorescence of the vitelline membrane (surrounding the entire embryo) as well as the excitation/emission light scattering within amnioserosa cells that undergo rapid morphological changes during development, which negatively impacts the structured illumination/emission profiles. Scale bar in (o) is 500 nm. Arrows in each image refer to the line of the shown cross-section. Experiments were repeated with >1000 images with similar results.

We should note that the aberrations or artifacts potentially observed in some of the ground truth training images can couple back into the network's inference and result in some residual artifacts in the network output. If the ground truth training image set is not dominated with such artifacts, the impact of this would be negligible, close to the noise floor of the output image. Such residual artifacts can be further reduced by pre-selection of the training ground truth images to be free from major artifacts (if possible) or through an additional loss term applied to suppress such features during the training process.

2.5 Depth-of-field enhancement

Another important feature of the deep network-based image transformation approach is that it can resolve features over an extended DOF because of the lower NA of the input image (**Figure 2.2**, **Figure 2.3**, and **Figure 2.5**). This phenomenon is further illustrated by acquiring a depth-resolved image set (composed of 34 images, axially-separated by 0.3 μm) corresponding to the blood-vessel sample using a 20×/0.75NA objective, and synthesized an extended-DOF image using the ImageJ plugin EDF, [125] which provides a significantly improved ground truth image compared to a single high-resolution image. These results and the comparison reported in **Figure 2.14** clearly demonstrate the extended-DOF capabilities of the super-resolution method. This extended DOF is also favorable in terms of photo-damage to the sample, by eliminating the need for a fine axial-scan within the sample volume, which might reduce the overall light delivered to the sample, while also making the imaging process more efficient. Although some thicker samples will ultimately require axial-scanning, the presented approach will still reduce the number of scans required by inferring high-resolution images from parts of the sample that would have been defocused when using higher NA imaging systems (**Figure 2.3**).

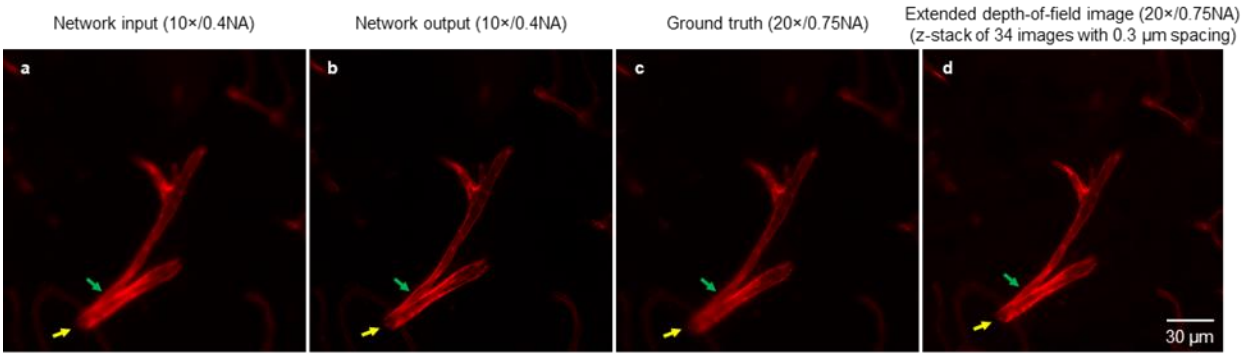

Figure 2.14 Demonstration of extended depth-of-focus (DOF) of the network with a mouse brain blood vessel sample. (a) The network input image captured with a 10×/0.4NA objective lens. (b) The network output image with a model pre-trained with only F-actin images. (c) High-resolution image captured with a 20×/0.75NA objective lens. The focusing plane is automatically selected by the microscope software using an auto-focusing algorithm. (d) An extended-DOF image synthesized from a z-stack of 34 high resolution images (separated axially by 0.3 μm) using ImageJ Plugin EDF [125]. The output image from a single input image is demonstrated and compared to extended-DOF image. Artifact analysis

A common concern for computational approaches that enhance image resolution is the potential emergence of spatial artifacts which may degrade the image quality, such as the Gibbs phenomenon in Lucy-Richardson deconvolution. [126] To explore this, I randomly selected an example in the test image dataset, and quantified the artifacts of the network output using the NanoJ-Squirrel Plugin [107]. The plugin iteratively estimates a resolution scaling function (RSF) from the low-resolution (LR) image to the high-resolution (HR) image, convolves the HR image with this RSF and calculates its pixel-wise absolute difference from the LR image. The plugin also provides two globally averaged scores: Resolution Scaled Error (RSE) and Resolution Scaled Pearson coefficient (RSP), defined as:

$$\mathrm{RSE}(f,g)=\sqrt{\frac{\sum_{x,y}\left(f(x,y)-g(x,y)\right)^2}{n}}$$
$$\mathrm{RSP}(f,g)=\frac{\sum_{x,y}\left(f(x,y)-\bar{f}\right)\left(g(x,y)-\bar{g}\right)}{\sqrt{\sum_{x,y}\left(f(x,y)-\bar{f}\right)}\sqrt{\sum_{x,y}\left(g(x,y)-\bar{g}\right)}} \tag{2.2}$$

where, f and g are the LR and simulated LR images, respectively, and ($\bar{}$) refers to the two-dimensional mean operator. Generally, the RSE is more sensitive to brightness and contrast differences, while the RSP helps to assess the image qualities across modalities, by quantifying their correlation.

In my implementation using this plugin, the "Reference image" was set to the LR input image, the "Super-resolution reconstruction" was set to the network output image. "RSF Estimate Image" was set to "RSF unknown, estimate via optimization" with "Max. Mag. in Optimization" set to 5. The error map of the network's output image with respect to the network's input (LR image) is shown in **Figure 2.15**g, resulting in RSE = 0.912 and RSP = 0.999.

I then repeated the same operations detailed above, estimating the error map between the low-resolution input image and the ground truth (HR) image, as shown in **Figure 2.15**j, which resulted in RSE = 1.509 and RSP = 0.998. These results show that the network output image does not generate noticeable super-resolution related artifacts and in fact has the same level of spatial mismatch error that the ground truth HR image has with respect to the LR input image (with a correlation of ~1 and an absolute error ~1 out of 255). This conclusion is further confirmed by **Figure 2.15**f, which overlays the network output image and the ground truth image in different colors, revealing no obvious feature mismatch between the two. The same conclusion remained consistent for other test images as well.

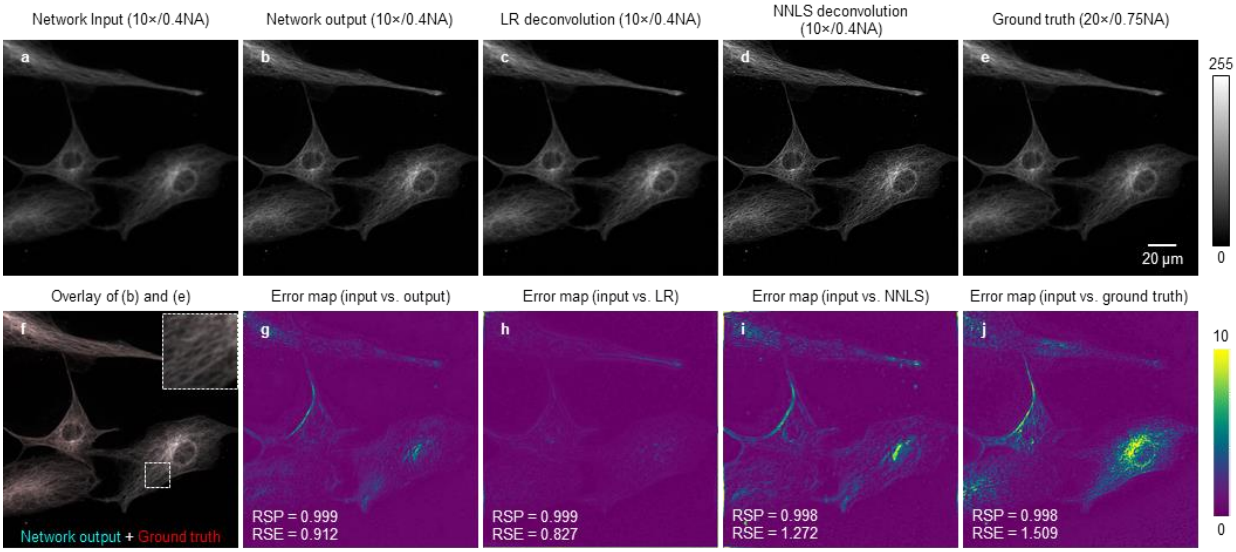

Figure 2.15 Quantification of super-resolution artifacts using the NanoJ-Squirrel Plugin. [127] (a) Network input, (b) network output, (c) Lucy-Richardson (LR) deconvolution, (d) non-negative least square (NNLS) deconvolution, and (e) ground truth images of the microtubule structure inside a BPAEC. (f) Overlay image of (b) in cyan and (e) in red shows sharp features without red or cyan color blocks, which means there is no obvious feature mismatch between the network output image and the ground truth image. (g-j) Error maps of the network input image vs. the network output image (g), LR deconvolution image (h), NNLS deconvolution image (j), and the ground truth image (j), calculated by NanoJ-Squirrel. All the maps (g-j) show high RSP (resolution scale Pearson-correlation) scores that are almost 1, and low resolution scaled error (RSE) scores of ~1, out of 255. Note that the network output image has better agreement in RSE than the ground truth image. This can be partially explained by the larger depth-of-field of a low NA objective lens that is used for acquiring the network input image. Analysis was performed on a randomly selected image from a group of 94 testing images with similar results.

2.6 Materials and methods

Wide-field fluorescence microscopic image acquisition

The fluorescence microscopic images (**Figure 2.1** and **Figure 2.2**) were captured by scanning a microscope slide containing multi-labeled bovine pulmonary artery endothelial cells (BPAEC) (FluoCells Prepared Slide #2, Thermo Fisher Scientific) on a standard inverted microscope which

is equipped with a motorized stage (IX83, Olympus Life Science). The low-resolution (LR) and high-resolution (HR) images were acquired using 10×/0.4NA (UPLSAPO10X2, Olympus Life Science) and 20×/0.75NA (UPLSAPO20X, Olympus Life Science) objective lenses, respectively. Three bandpass optical filter sets were used to image the three different labelled cell structures and organelles: Texas Red for F-actin (OSFI3-TXRED-4040C, EX562/40, EM624/40, DM593, Semrock), FITC for microtubules (OSFI3-FITC-2024B, EX485/20, EM522/24, DM506, Semrock), and DAPI for cell nuclei (OSFI3-DAPI-5060C, EX377/50, EM447/60, DM409, Semrock). The imaging experiments were controlled by MetaMorph microscope automation software (Molecular Devices), which performed translational scanning and auto-focusing at each position of the stage. The auto-focusing was performed on the FITC channel, and the DAPI and Texas Red channels were both exposed at the same plane as FITC. With a 130 W fluorescence light source set to 25% output power (U-HGLGPS, Olympus Life Science), the exposure time for each channel was set to: Texas Red 350 ms (10×) and 150 ms (20×), FITC 800 ms (10×) and 400 ms (20×), DAPI 60 ms (10×) and 50 ms (20×). The images were recorded by a monochrome scientific CMOS camera (ORCA-flash4.0 v2, Hamamatsu Photonics K.K.) and saved as 16-bit grayscale images with regards to each optical filter set. The additional test images (**Figure 2.3**) were captured using the same setup with FluoCells Prepared Slide #1 (Thermo Fisher Scientific), with the filter setting of Texas Red for mitochondria, FITC for F-actin, and FluoCells Prepared Slide #3 (Thermo Fisher Scientific), with the filter setting of Texas Red for actin, and FITC for glomeruli and convoluted tubules. The mouse brain tumour sample was prepared with mouse brains perfused with Dylight 594 conjugated Tomato Lectin (1 mg/ml) (Vector Laboratories, CA), fixed in 4% para-formaldehyde for 24 hours and incubated in 30% sucrose in phosphate-buffered

saline, then cut in 50 μm thick sections as detailed in [[128]], and imaged using Texas Red filter set for blood vessels, and FITC filter set for tumour cells.

Confocal and STED image acquisition

For the Histone 3 imaging experiments, the HeLa cells were grown as a monolayer on high-performance coverslips (170 μm +/- 10 μm) and fixed with methanol. Nuclei were labelled with a primary Rabbit anti-Histone H3 trimethyl Lys4 (H3K4me3) antibody (Active motif # 39159) and a secondary Atto-647N Goat anti-rabbit IgG antibody (Active Motif # 15048) using the reagents of the MAXpack Immunostaining Media Kit (Active Motif # 15251). The labelled cells were then embedded with Mowiol 4-88 and mounted on a standard microscope slide.

The nano-bead samples for confocal and STED experiments (**Figure 2.6** and **Figure 2.7**) were prepared with 20 nm fluorescent nano-beads (FluoSpheres Carboxylate-Modified Microspheres, crimson fluorescent (625/645), 2% solids, Thermo Fisher Scientific) that were diluted 100 times with methanol and sonicated for 3×10 minutes, and then mounted with antifade reagents (ProLong Diamond, Thermo Fisher Scientific) on a standard glass slide, followed by placing on high-performance coverslips (170 μm +/- 10 μm) (Carl Zeiss Microscopy).

Samples were imaged on a Leica TCS SP8 STED confocal using a Leica HC PL APO 100×/1.40 Oil STED White objective. The scanning for each FOV was performed by a resonant scanner working at 8000 Hz with 16 times line average and 30 times frame average for nanobeads, and 8 times line average and 6 times frame average for cell nuclei. The fluorescent nano-beads were excited with a laser beam at 633 nm wavelength. The emission signal was captured with a hybrid photodetector (HyD SMD, Leica Microsystems) through a 645~752 nm bandpass filter. The excitation laser power was set to 5% for confocal imaging, and 50% for STED imaging, so

that the signal intensities remained similar while keeping the same scanning speed and gain voltage. A depletion beam of 775 nm was also applied when capturing STED images with 100% power. The confocal pinhole was set to 1 Airy unit (e.g., 168.6 µm for 645 nm emission wavelength and 100× magnification) for both the confocal and STED imaging experiments. The cell nuclei samples were excited with a laser beam at 635 nm and captured with the same photodetector which is set to 1× gain for confocal and 1.9× gain for STED with a 650-720 nm bandpass filter. The confocal pinhole was set to 75.8 µm (e.g., 0.457 Airy unit for 650 nm emission wavelength and 100× magnification) for both the confocal and STED imaging experiments. The excitation laser power was set to 3% and 10% for confocal and STED experiments, respectively. The scanning step size (i.e., the effective pixel size) for both experiments was ~30 nm to ensure sufficient sampling rate. All the images were exported and saved as 8-bit grayscale images.

TIRF-SIM image acquisition

Gene edited SUM159 cells expressing AP2-eGFP [129] were grown in F-12 medium containing hydrocortisone, penicillin-streptomycin and 5% fetal bovine serum (FBS). Transient expression of mRuby-CLTB (Addgene; Plasmid #55852) was carried using Gene Pulser Xcell electroporation system (Bio-Rad Laboratories, CA, USA) following the manufacturer's instructions, and imaging was performed 24-48 hours after transfection. Cells were imaged in phenol-red-free L15 (Thermo Fisher Scientific) supplemented with 5% FBS at 37°C ambient temperature. Clathrin dynamics were monitored in lateral epidermis and amnioserosa tissues of Drosophila embryos using UAS/GAL4 system as described in Ref. [[130]]. Drosophila embryos were gently pressed against the cover glass to position the apical surface of the lateral epidermis and amnioserosa cells within the evanescence field of the TIRF system. Arm-GAL4 strain was provided by the Bloomington Drosophila Stock Center; CLC-mEmerald strain was provided by

Dr. Henry Chang (Purdue University, USA). TIRF-SIM images were acquired by a 100×/1.49NA objective lens (Olympus Life Science, CA, USA) fitted on an inverted microscope (Axio Observer; ZEISS) equipped with a sCMOS camera (ORCA-Flash4.0; Hamamatsu). Structured illumination was provided by a spatial light modulator as described in Ref. [[111]].

Image pre-processing

For wide-field images (**Figure 2.1**, **Figure 2.2**, and **Figure 2.10**), a low intensity threshold was applied to subtract background noise and auto-fluorescence, as a common practice in fluorescence microscopy. The threshold value was estimated from the mean intensity value of a region without objects, which is ~300 out of 65535 in the 16-bit images. The LR images are then linearly interpolated two times to match the effective pixel size of the HR images. Accurate registration of the corresponding LR and HR training image pairs is of crucial importance since the objective function of the network consists of adversarial loss and pixel-wise loss. I employed a two-step registration workflow to achieve the needed registration with sub-pixel level accuracy. First, the fields-of-view of LR and HR images are digitally stitched in a MATLAB script interfaced with Fiji [131] Grid/Collection stitching plugin [132] through MIJ [133], and matched by fitting their normalized cross-correlation map to a 2D Gaussian function and finding the peak location. However, due to optical distortions and color aberrations of different objective lenses, the local features might still not be exactly matched. To address this, the globally matched images are fed into a pyramidal elastic registration algorithm to achieve sub-pixel level matching accuracy, which is an iterative version of the registration module in Fiji Plugin NanoJ, with a shrinking block size. [6,107,131,134] This registration step starts with a block size of 256×256 and stops at a block size of 64×64, while shrinking the block size by 1.2 times every 5 iterations with a shift tolerance of 0.2 pixels. Due to the slightly different placement and the distortion of the optical filter sets, I

performed the pyramidal elastic registration for each fluorescence channel independently. At the last step, the precisely registered images were cropped 10 pixels on each side to avoid registration artifacts, and converted to single-precision floating data type and scaled to a dynamic range of 0~255. This scaling step is not mandatory but creates convenience for fine tuning of hyper-parameters when working with images from different microscopes/sources.

For confocal and STED images (**Figure 2.6**, **Figure 2.7**, and **Figure 2.9**) which were scanned in sequence on the same platform, only a drift correction step was required, which was calculated from the 2D Gaussian fit of the cross-correlation map. The drift was found to be ~10 nm for each scanning FOV between the confocal and STED images. Thresholding was not performed to the nanobead dataset for the network training. However, after the test images were enhanced by the network, I subtracted a constant value (calculated by taking the mean value of an empty region) from the confocal (network input), the super-resolved (network output), and the STED (ground truth) images, respectively, for better visualization and comparison of the images. The total number of images used for training, validation and blind testing of each network are summarized in **Table 2.1**.

Table 2.1 Number of experimental image datasets used for each network. Each image has 1024×1024 pixels.

Super-resolution network	Number of training image pairs	Number of validation image pairs	Number of testing image pairs
Wide-field (TxRed)	1945	680	94
Wide-field (FITC)	1945	680	94
Wide-field (DAPI)	1945	680	94
Confocal-STED (nanobeads)	607	75	75
Confocal-STED (transfer learning)	1100	100	30
TIRF-SIM	3003	370	1100

Generative adversarial network structure and training

In this work, the deep neural network was trained following the generative adversarial network (GAN) framework [57], which has two sub-networks being trained simultaneously, a generative model which enhances the input LR image, and a discriminative model which returns an adversarial loss to the resolution-enhanced image, as illustrated in **Figure 2.16**. I designed the objective function as the combination of the adversarial loss with two regularization terms: the mean square error (MSE), and the structural similarity (SSIM) index [135]. Specifically, the optimizers aim to minimize:

$$\begin{aligned}\mathcal{L}(G;D) &= -\log D\big(G(x)\big) + \lambda \times \mathrm{MSE}\big(G(x), y\big) - \nu \times \log\Big[\big(1+\mathrm{SSIM}\big(G(x), y\big)\big)/2\Big] \\ \mathcal{L}(D;G) &= -\log D(y) - \log\Big[1 - D\big(G(x)\big)\Big]\end{aligned} \tag{2.3}$$

where x is the LR input, $G(x)$ is the generative model output, $D(\cdot)$is the discriminative model prediction of an image (network output or ground truth image), and y is the HR image used as ground truth. The structural similarity index is defined as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{2.4}$$

where μ_x, μ_y are the averages of x, y; σ_x^2, σ_y^2 are the variances of x, y; $\sigma_{x,y}$ is the covariance of x and y; and c_1, c_2 are the variables used to stabilize the division with a small denominator. An SSIM value of 1.0 refers to identical images. When training with the wide-field fluorescence images, the regularization constants λ and ν were set to accommodate the MSE loss and the SSIM loss to be ~1%-10% of the combined generative model loss $\mathcal{L}(G;D)$, depending on the noise level of the image dataset. When training with the confocal-STED image datasets, λ was kept the same and set ν to 0. While the adversarial loss guides the generative model to map the LR images into HR,

the two regularization terms assure that the generator output image is established on the input image with matched intensity profile and structural features. These two regularization terms also help us stabilize the training schedule and smoothen out the spikes on the training loss curve before it reaches equilibrium. For the sub-network models, a similar network structure was employed as described in Ref. [[134]]. The relatively low weight that is given to the MSE and SSIM terms is due to the fact that these values already represent a high degree of agreement between the low-resolution input and the gold standard label (for example, ~ 0.87 – 0.94 for the wide-field microscopy experiments). Hence, a large weight given to these loss terms will drive the network to converge to a local minimum that will strongly resemble the low-resolution input and not learn the desired (super-resolved) output distribution. Therefore, it might be beneficial for some other applications to increase the weights of these terms, e.g., for low SNR images, where the task of denoising might be of main interest for automated segmentation and related image processing tasks.

Generative Model

U-net is a CNN architecture, which was first proposed for medical image segmentation, yielding high performance with very few training datasets. [56] A similar network architecture has also been successfully applied in recent image reconstruction and virtual staining applications [13,134]. The structure of the generative network used in this work is illustrated in **Figure 2.16**, which consists of four down-sampling blocks and four up-sampling blocks. Each down-sampling block consists of three residual convolutional blocks, within which it performs:

$$x_k = x_{k-1} + \mathrm{LReLU}[\mathrm{Conv}\{\mathrm{LReLU}[\mathrm{Conv}\{\mathrm{LReLU}[\mathrm{Conv}\{x_{k-1}\}]\}]\}], k = 1,2,3,4 \qquad (2.5)$$

where x_k represents the output of the *k*-th down-sampling block, and x_0 is the LR input image. Conv{ } is the convolution operation, LReLU[ ] is the leaky rectified linear unit activation function with a slope of $\alpha = 0.1$, i.e.,

$$\mathrm{LReLU}(x;\alpha) = \mathrm{Max}(0,x) - \alpha \times \mathrm{Max}(0,-x) \tag{2.6}$$

The input of each down-sampling block is zero-padded and added to the output of the same block. The spatial down-sampling is achieved by an average pooling layer after each down-sampling block. A convolutional layer lies at the bottom of this U-shape structure that connects the down-sampling and up-sampling blocks.

Each up-sampling block also consists of three convolutional blocks, within which it performs:

$$y_k = \mathrm{LReLU}[\mathrm{Conv}\{\mathrm{LReLU}[\mathrm{Conv}\{\mathrm{LReLU}[\mathrm{Conv}\{\mathrm{Concat}(x_{5-k}, y_{k-1})\}]\}]\}], k = 1,2,3,4 \tag{2.7}$$

where y_k represents the output of the *k*-th up-sampling block, and y_0 is the input of the first up-sampling block. $\mathrm{Concat}(\,)$ is the concatenation operation of the down-sampling block output and the up-sampling block input on the same level in the U-shape structure. The last layer is another convolutional layer that maps the 32 channels into 1 channel that corresponds to a monochrome grayscale image.

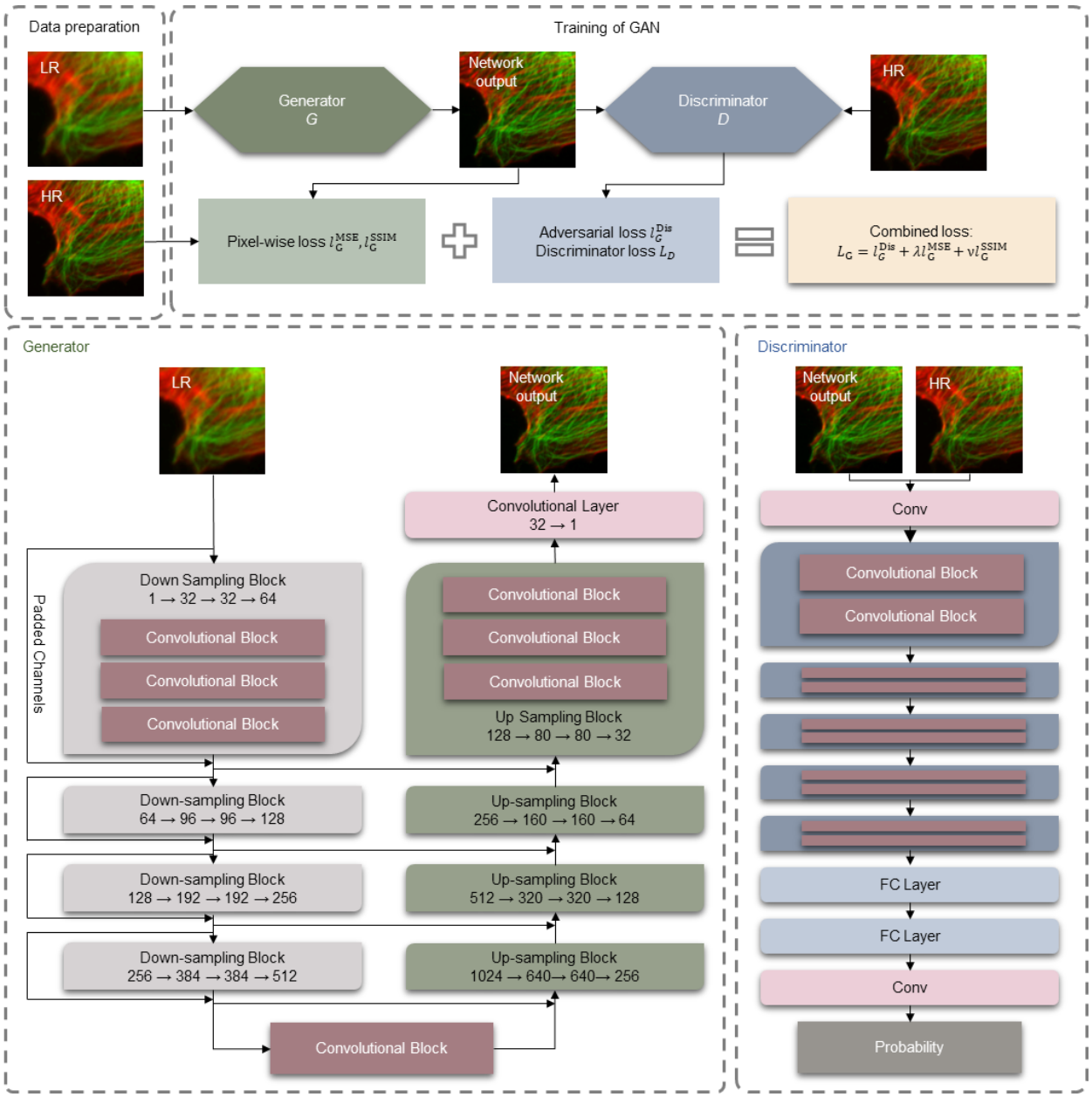

Figure 2.16 The training process and the architecture of the generative adversarial network (GAN) that were used for image super-resolution.

Discriminative Model

As shown in **Figure 2.16**, the structure of the discriminative model begins with a convolutional layer, which is followed by 5 convolutional blocks, each of which performs the following operation:

$$z_k = \mathrm{LReLU}[\mathrm{Conv}\{\mathrm{LReLU}[\mathrm{Conv}\{z_{k-1}\}]\}], k = 1,2,3,4,5 \quad (2.8)$$

Where z_k represents the output of the *k*-th convolutional block, and z_0 is the input of the first convolutional block. The output of the last convolutional block is fed into an average pooling layer whose filter shape is the same as the patch size, i.e., $H \times W$. This layer is followed by two fully connected layers for dimension reduction. The last layer is a sigmoid activation function whose output is the probability of an input image being ground truth, defined as:

$$D(z) = \frac{1}{1+\exp(-z)} \quad (2.9)$$

Network training schedule

During the network training the patch size is set to be 64×64, with a batch size of 12 on each of the two GPUs. Within each iteration, the generative model and the discriminative model are each updated once while keeping the other unchanged. Both the generative model and the discriminative model were randomly initialized and optimized using the adaptive moment estimation (Adam) optimizer [136] with a starting learning rate of 1×10^{-4} and 1×10^{-5}, respectively. This framework was implemented with TensorFlow framework version 1.7.0 [137] and Python version 3.6.4 in Microsoft Windows 10 operating system. The training was performed on a consumer grade laptop (EON17-SLX, Origin PC) equipped with dual GeForce GTX1080 graphic cards (NVDIA) and a Core i7-8700K CPU @ 3.7GHz (Intel). The final model for wide-field images were selected with the smallest validation loss at around ~50,000th iteration, which took ~10 hours to train. The final model for confocal-STED transformation (**Figure 2.6**, **Figure 2.7**) is selected with the smallest validation loss at around ~500,000th iteration, which took ~90 hours to train. The transfer learning for confocal-STED transformation network (**Figure 2.9**) was

implemented with the same framework on a desktop computer with dual GTX1080Ti graphic cards, while setting the patch size to be 256 × 256 with 4 patches on each GPU. It was first initialized with confocal-STED model trained with nano-beads, and then refined with cell nuclei image data with ~20,000 iterations, which took ~24 hours. The training of TIRF to TIRF-SIM transformation network was also implemented with dual GTX1080Ti graphic cards, while setting the patch size to be 64×64, and 64 patches on each GPU. The final model was trained for ~20,000 iterations which took ~18 hours.

Implementation of LR and NNLS deconvolution

To make a fair comparison, the lower resolution images were up-sampled 2 times by bilinear interpolation before being deconvolved. The Born and Wolf PSF model [138,139] was used with parameters set to match the experimental setup, i.e., NA = 0.4, immersion refractive index = 1.0, pixel size = 325 nm. The PSF is generated by an Fiji PSF Generator Plugin [131,140]. An exhaustive parameter search was performed by running the Lucy-Richardson algorithm with 1~100 iterations and damping threshold 0%~10%. The results were visually assessed, with the best one obtained at 10 iterations and 0.1% damping threshold (**Figure 2.2**, third column). The NNLS deconvolution was performed with Fiji Plugin DeconvolutionLab2 [141] with 100 iterations and a step size of 0.5. The deconvolution for Texas Red, FITC, and DAPI channels were performed separately, assuming the central emission wavelengths to be 630 nm, 532nm, and 450 nm, respectively.

Characterization of the lateral resolution by PSF fitting

The resolution differences among the network input (confocal), the network output (confocal), and the ground truth (STED) images were characterized by fitting their PSFs to a 2D Gaussian

profile, as shown in **Figure 2.7**. To do so, more than 400 independent bright spots were selected from the ground truth STED images and cropped out with the surrounding 19×19-pixel regions, i.e., ~577×577 nm^2. The same locations were also projected to the network input and output images, followed by cropping of the same image regions as in the ground truth STED images. Each cropped region was then fitted to a 2D Gaussian profile. The FWHM values of all these 2D profiles were plotted as histograms, shown in **Figure 2.7**. For each category of images, the histogram profile within the main peak region is fitted to a 1D Gaussian function (**Figure 2.7**). A similar process was repeated for the results reported in **Figure 2.5**d.

2.7 Discussion

The deep learning approach allows for the generation of super-resolution images directly from images acquired on conventional, diffraction-limited microscopes without *a priori* knowledge about the sample and/or the image formation process. In addition to democratizing super-resolution microscopy, this approach offers the benefits of rapidly imaging larger fields-of-view and depths-of-field, creating higher resolution images with fewer frames and/or lower light doses, which enables new opportunities for imaging objects with reduced photo-bleaching and photo-toxicity. [117,120]

An essential step of the presented super-resolution framework is the accurate alignment and registration between the lower resolution and the higher resolution label images. This multi-stage image registration process (see 2.6 Materials and methods**)** allows the network to learn a pixel-to-pixel transformation and is used as a regularization for the network to learn the resolution enhancement, while avoiding warping of the input images, which in turn significantly reduces potential artifacts.

The deep learning approach also improves the image SNR. In fact, the resolution limit of a microscopy modality is fundamentally limited by its SNR [142]; stated differently, the lack of some spatial frequencies at the image plane (e.g., carried by evanescent waves) does not pose a fundamental limit for the achievable resolution of a computational microscope. These missing spatial frequencies (although not detected at the image) can in principle be extrapolated based on the measured or known spatial frequencies of an object. [142] For example, the full spatial frequency spectrum of an object function that has a limited spatial-extent with finite energy can in theory be recovered from the partial knowledge of its spectrum using the analytical continuation principle since its Fourier transform defines an entire function. [143] In practice, however, this is a challenging task and the success of such a frequency extrapolation method is strongly dependent on the SNR of the measured image information and *a priori* information regarding the object. Although the presented neural network-based super-resolution approach does not include any such analytical continuation models, or any *a priori* assumptions about the known frequency bands or support information of the object, through image data it learns to statistically separate out noise patterns from the structural information of the object, helping us achieve effectively much improved frequency extrapolation and resolution enhancement compared to the state-of-the-art methods as reported in the **Results**.

To practice this approach on new types of samples or new imaging systems that were not part of the training process, fresh application of this presented framework is recommended for getting optimal results, starting with the image registration between the input images (lower resolution) and the desired labels (higher resolution), followed by the training of a GAN. Transfer learning from a previously trained network for another type of sample might speed up the convergence of this learning process; however, this is neither a required step nor a replacement for the entire image

registration and GAN training processes performed on new sample types of interest. After a sufficiently large number of training iterations (e.g., >10,000) the optimal network model can be selected when the validation loss value no longer decreases.

Taken together, this work represents an important step forward for the fields of computational microscopy and super-resolution imaging, and should help us democratize high-resolution imaging systems, potentially enabling new biological observations beyond what can be achieved in well-resourced institutions and laboratory settings. The ability to close the gap between lower resolution and higher resolution imaging systems using deep learning framework is fundamentally tied to image SNR in both the training and blind testing phases, and in this sense the presented image transformation framework is limited in its performance by noise, very much like all the other super-resolution imaging modalities.

Chapter 3 Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning

3.1 Introduction

Microscopic imaging of tissue samples is a fundamental tool used for the diagnosis of various diseases and forms the workhorse of pathology and biological sciences. The clinically-established gold standard image of a tissue section is the result of a laborious process, which includes the tissue specimen being formalin-fixed paraffin-embedded (FFPE), sectioned to thin slices (typically ~2-10 μm), labeled/stained and mounted on a glass slide, which is then followed by its microscopic imaging using e.g., a bright-field microscope. All these steps use multiple reagents and introduce irreversible effects on the tissue. There have been recent efforts to change this workflow using different imaging modalities. One line of work imaged fresh, non-paraffin-embedded tissue samples using non-linear microscopy methods based on e.g., two-photon fluorescence, second harmonic generation [144], third-harmonic generation [145] as well as Raman scattering [146–148]. Another study used a controllable super-continuum source [149] to acquire multi-modal images for chemical analysis of fresh tissue samples. These methods require using ultra-fast lasers or super-continuum sources, which might not be readily available in most settings and require relatively long scanning times due to weaker optical signals. In addition to these, other microscopy methods for imaging non-sectioned tissue samples have also emerged by using UV-excitation on stained samples [150,151], or by taking advantage of the auto-fluorescence emission of biological tissue at short wavelengths [152]. In fact, there are unique opportunities using auto-fluorescence for imaging tissue samples by making use of the fluorescent light emitted from endogenous fluorophores. It has been demonstrated that such endogenous fluorescence signatures carry useful information that can be mapped to functional and structural properties of biological specimen and

therefore have been used extensively for diagnostics and research purposes [152–154]. One of the main focus areas of these efforts has been the spectroscopic investigation of the relationship between different biological molecules and their structural properties under different conditions. Some of these well characterized biological constitutes include vitamins (e.g., vitamin A, riboflavin, thiamin), collagen, coenzymes, fatty acids, among others [153].

While some of the above discussed techniques have unique capabilities to discriminate e.g., cell types and sub-cellular components in tissue samples using various contrast mechanisms, pathologists as well as tumor classification software [155] are in general trained for examining histologically stained tissue samples to make diagnostic decisions. Partially motivated by this, some of the above mentioned techniques were also augmented to create pseudo-Hematoxylin and Eosin (H&E) images [144,156], which were based on a linear approximation that relates the fluorescence intensity of an image to the dye concentration per tissue volume, using empirically determined constants that represent the mean spectral response of various dyes embedded in the tissue. These methods also used exogenous staining to enhance the fluorescence signal contrast in order to create virtual H&E images of tissue samples.

The deep learning-based cross-modality image transformation shows a unique opportunity here: building a non-physics-based transformation from auto-fluorescence image of unlabelled samples to bright-field images of histochemically stained samples. Part of this chapter has been previously published in:

- Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nat. Biomed. Eng. 3, 466 (2019).

In this work, deep learning-based virtual histology staining is demonstrate using auto-fluorescence of unstained tissue, imaged with a wide-field fluorescence microscope through a standard near-UV excitation/emission filter set (see the 3.5 Methods section). The virtual staining is performed on a *single* auto-fluorescence image of the sample by using a deep Convolutional Neural Network (CNN), which is trained using the concept of Generative Adversarial Networks (GAN) [157] to match the bright-field microscopic images of tissue samples after they are labeled with a certain histology stain (**Figure 3.1**, **Figure 3.2**). Therefore, using a CNN, the histological staining and bright-field imaging steps are replaced with the output of the trained neural net, which is fed with the auto-fluorescence image of the unstained tissue. The network inference is fast, taking e.g., 1.9 seconds/mm^2 using a desktop computer for a tissue section scanned using a 20× objective lens and can be significantly improved by using ever evolving computing hardware with parallelization capabilities.

This deep learning-based virtual histology staining method was demonstrated by imaging label-free human tissue samples including salivary gland, thyroid, kidney, liver and lung, where the network output created equivalent images, very well matching the images of the same samples that were labeled with three different stains, i.e., H&E (salivary gland and thyroid), Jones stain (kidney) and Masson's Trichrome (liver and lung). Furthermore, the staining efficacy of this approach for whole slide images (WSIs) corresponding to some of these samples was blindly evaluated by a group of pathologists, who were able to recognize histopathological features with the virtual staining technique, achieving a high degree of agreement with the histologically stained images of the same samples, as will be detailed in the Results section.

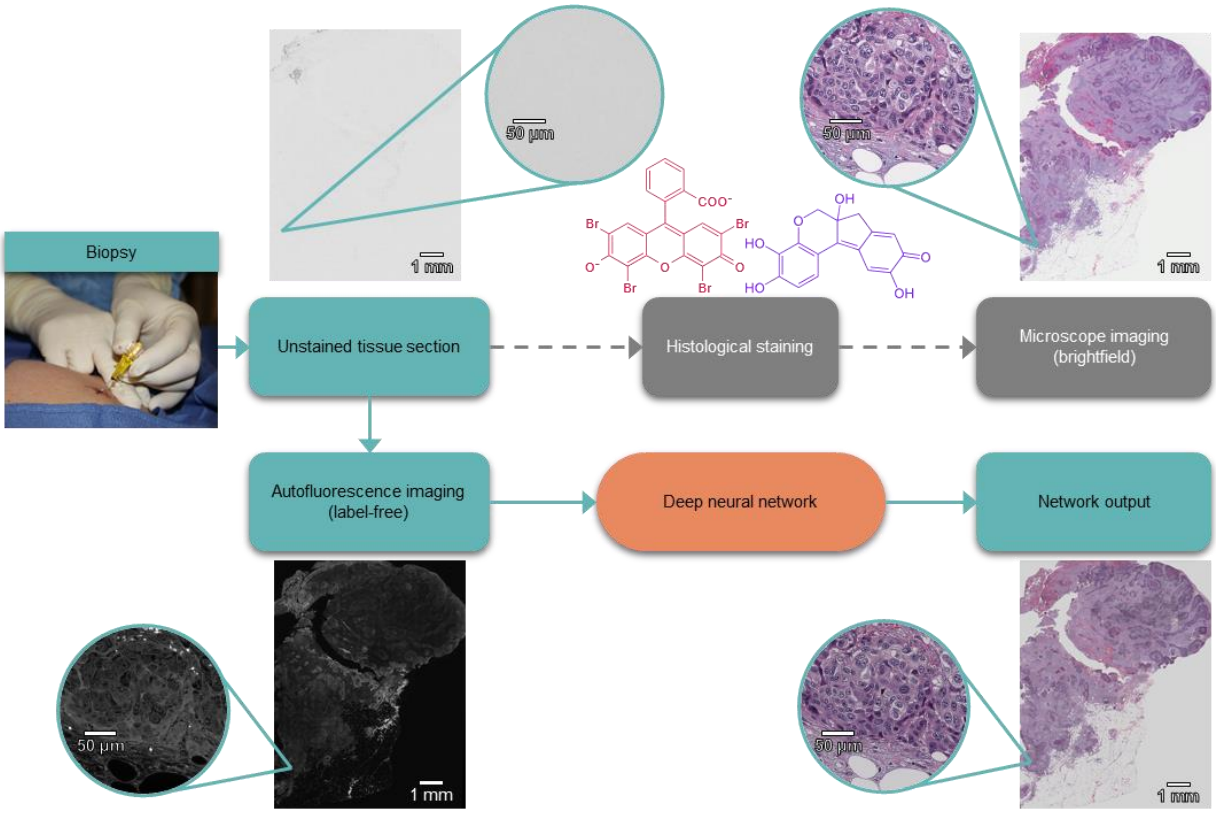

Figure 3.1 Deep-learning-based virtual histology staining using autofluorescence of unstained tissue. The schematic outlines the steps in the standard (top) and virtual (bottom) staining techniques. After training using a GAN, the neural network (orange box) rapidly outputs a virtually stained tissue image (H&E in this case), in response to the input of an autofluorescence image of an unstained tissue section, bypassing the standard histological staining procedure (grey boxes).

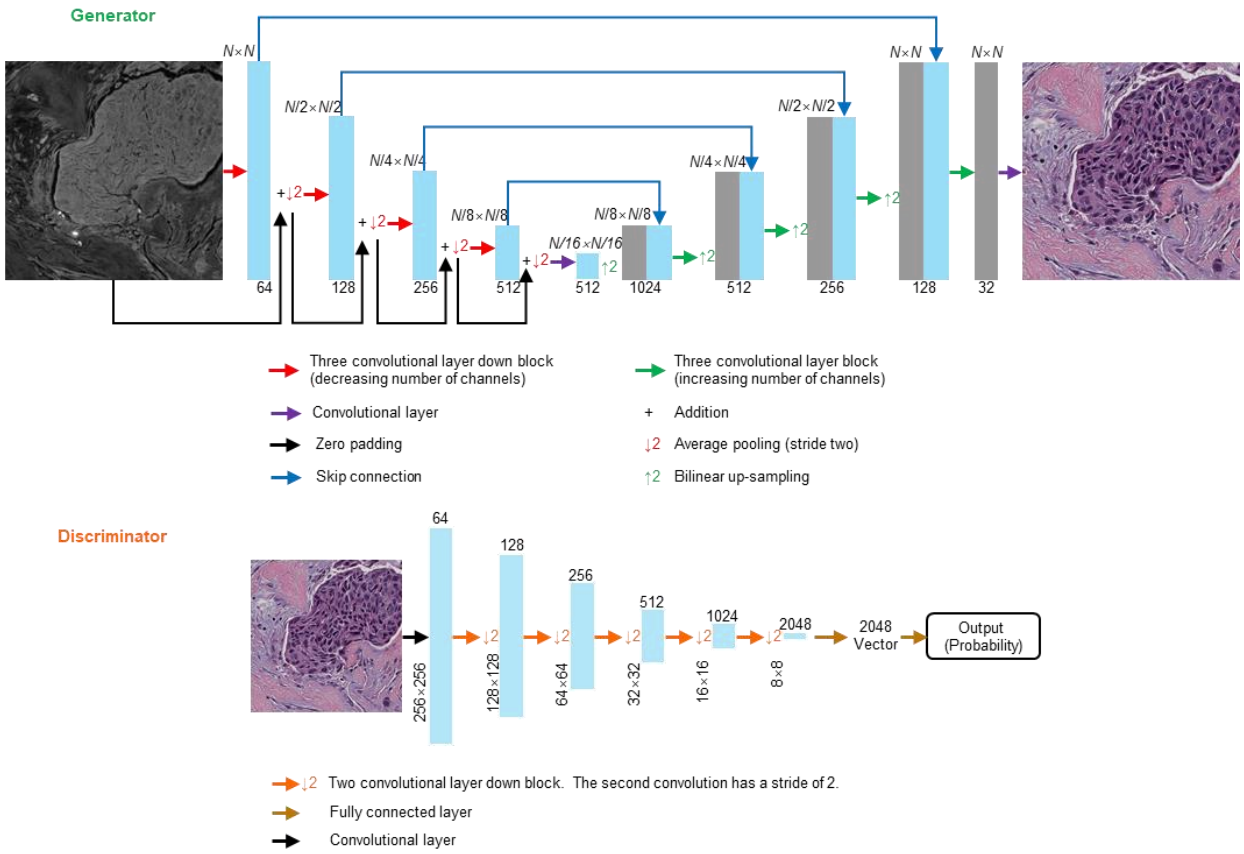

Figure 3.2 Virtual staining GAN architecture. Schematic of the CNN operation. The generator section is used to virtually stain the images. It comprises four 'down blocks', each of which are made up of three convolutional layers that are each followed by an average pooling layer of stride two. The down section is followed by four 'up blocks', which each contain three convolutional layers and are bilinearly upsampled by a factor of two. Skip connections are used to pass data between layers of the same level. The discriminator comprises five down blocks, each of which has two convolutional layers; the second convolutional layer has a stride of two, to reduce the tensor size. The down block reduces the size of the images while increasing the number of channels and is followed by two fully connected layers. The variable n represents the number of pixels on the lateral dimensions of each image patch that passes through the network. During the training, a 256 × 256 pixel patch is used; however, during the testing phase larger images can be inferred, as a result of the convolutional nature of the network.

Since the network's input image is captured by a conventional fluorescence microscope with a standard filter set, this approach has transformative potential to use unstained tissue samples for

pathology and histology applications, entirely bypassing the histological staining process, saving time and cost. For example, for the histology stains that were learned to virtually stain in this work, each staining procedure of a tissue section on average takes ~45 min (H&E) and 2-3 hours (Masson's Trichrome and Jones stain), with an estimated cost, including labor, of $2-5 [158,159] (H&E) and >$16-35 [159,160] (Masson's Trichrome and Jones stain). Furthermore, some of these histological staining processes require time-sensitive steps, demanding the expert to monitor the process under a microscope, which makes the entire process not only lengthy and relatively costly, but also laborious. The presented method bypasses all these staining steps, and also allows the preservation of unlabeled tissue sections for later analysis, such as micro-marking of sub-regions of interest on the unstained tissue specimen that can be used for more advanced immunohistochemical and molecular analysis to facilitate e.g., customized therapies [161,162]. Also note that, this deep learning-based virtual histology staining framework can be broadly applied to other excitation wavelengths or fluorescence filter sets, as well as to other microscopy modalities (such as non-linear microscopy) that utilize additional endogenous or exogenous contrast mechanisms [144–151]. In the experiments in this work, used sectioned and fixed tissue samples were used to be able to provide meaningful comparisons to the results of the standard histological staining process. However, the presented approach can potentially be applicable to non-fixed, non-sectioned tissue samples, potentially making it applicable to use in surgery rooms or at the site of a biopsy for rapid diagnosis or telepathology applications. Beyond its clinical applications, this method could broadly benefit histology field and its applications in life science research and education.

3.2 Virtual staining of tissue samples

The presented method was demonstrated using different combinations of tissue sections and stains. Following the training of a deep CNN (outlined in 3.5 Material and methods) I blindly tested its inference by feeding it with the auto-fluorescence images of label-free tissue sections that did not overlap with the images that were used in the training or validation sets. **Figure 3.3** summarizes the results for a salivary gland tissue section, which was virtually stained to match H&E stained bright-field images of the same sample. These results demonstrate the capability of the presented framework to transform an auto-fluorescence image of a label-free tissue section into a bright-field equivalent image, showing the correct color scheme that is expected from an H&E stained tissue. Evaluation of both **Figure 3.3**c and d show the H&E stains demonstrate a small island of infiltrating tumor cells within subcutaneous fibroadipose tissue. Note the nuclear detail, including distinction of nucleoli (arrow) and chromatin texture, is clearly appreciated in both panels. Similarly, in **Figure 3.3**g and h, the H&E stains demonstrate infiltrating squamous cell carcinoma. The desmoplastic reaction with edematous myxoid change (asterisk) in the adjacent stroma is clearly identifiable in both stains.

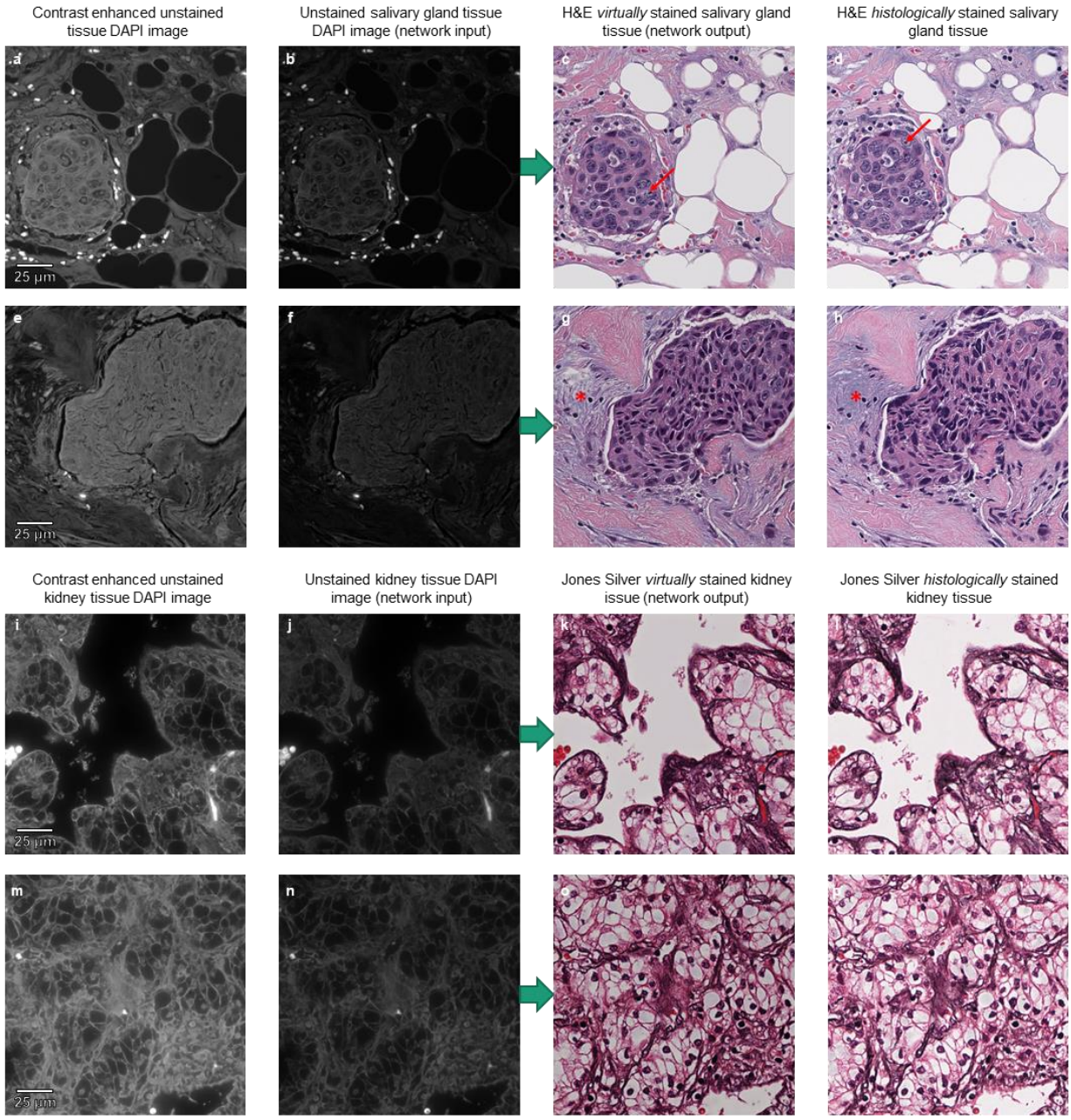

Figure 3.3 Virtual staining results match the Masson's trichrome stain for liver and lung tissue sections. **a**–**d**, Liver and lung tissue samples that are unstained (**a**,**b**), or either virtually (**c**) or histologically stained (**d**) with Masson's trichrome. **a**,**b**, Autofluorescence images of unstained liver tissue sections and unstained lung tissue sections. Only the raw images in **b** were used as input to the trained neural network. **c**, Virtual Masson's trichrome staining results (network output) for the same liver and lung tissue samples. **d**, Bright-field images of the same liver and lung tissue sections, after the histological staining process. Green arrows indicate the virtual staining of individual samples by the neural network.

Next, a deep network was trained to virtually stain other tissue types with two different "special" stains, i.e., the Jones methenamine silver stain (kidney) and the Masson's trichrome stain (liver and lung). **Figure 3.3** and **Figure 3.4** summarize the results for deep learning-based virtual staining of these tissue sections, which match very well to the bright-field images of the same samples, captured after the histological staining process. These results illustrate that the deep network can infer the staining patterns of different types of histology stains used for different tissue types, from a single auto-fluorescence mage of a label-free specimen. In **Figure 3.3**k,o the virtual Jones methenamine silver stain captures the black staining of extracellular collagen and maintains the visual integrity of the H&E counterstain in this example of renal cell carcinoma. The virtual Masson's trichrome staining in **Figure 3.4**c,g correctly reveals the histological features corresponding to hepatocytes, sinusoidal spaces, collagen and fat droplets (**Figure 3.4**g), consistent with the histologic appearance in the bright-field images of the same tissue samples, captured after the histological staining (**Figure 3.4**d,h). Similarly, the virtual staining of lung samples in **Figure 3.4**k,o reveals consistently stained histological features corresponding to vessels, collagen and alveolar spaces as they appear in the bright-field images after the histological staining (**Figure 3.4**l,p).

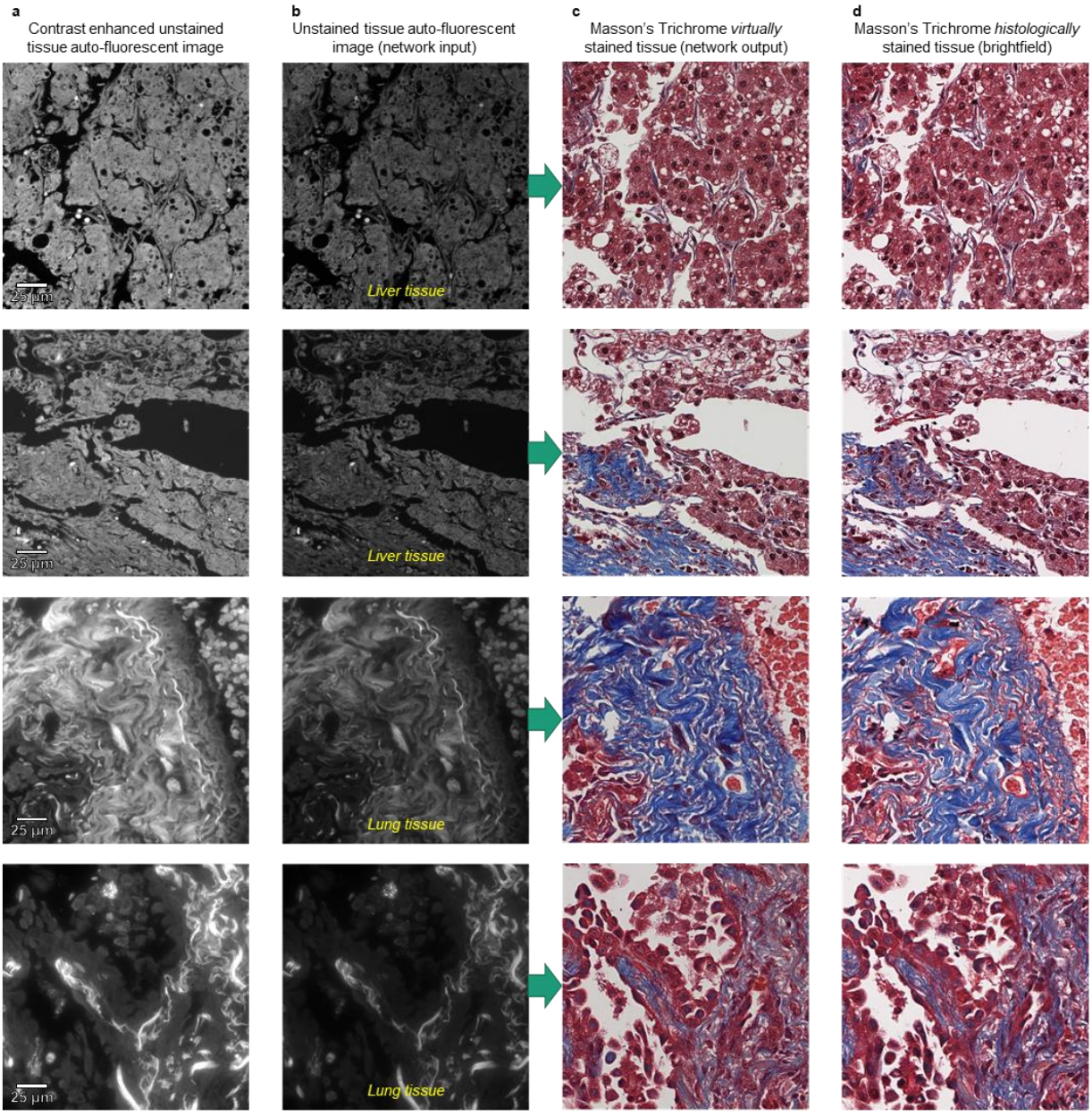

Figure 3.4 Virtual staining results match the Masson's trichrome stain for liver and lung tissue sections. **a**–**d**, Liver and lung tissue samples that are unstained (**a**,**b**), or either virtually (**c**) or histologically stained (**d**) with Masson's trichrome. **a**,**b**, Autofluorescence images of unstained liver tissue sections and unstained lung tissue sections. Only the raw images in **b** were used as input to the trained neural network. **c**, Virtual Masson's trichrome staining results (network output) for the same liver and lung tissue samples. **d**, Bright-field images of the same liver and lung tissue sections, after the histological staining process. Green arrows indicate the virtual staining of individual samples by the neural network.

The virtual staining framework was further compared to the standard histochemical staining for diagnosing multiple types of conditions on multiple types of tissues, which were either FFPE or frozen sections. The results are summarized in **Table 3.1**. The analysis of 15 tissue sections by four board-certified pathologists (who were not aware of our virtual staining technique) demonstrated 100% non-major discordance, which is defined as no clinically significant differences in diagnosis among professional observers. The 'time to diagnosis' varied considerably among observers, ranging from an average of 10 s per image for observer 2 to 276 s per image for observer 3. However, the intra-observer variability was very minor and tended towards a shorter time to diagnosis with our virtual stained slides for all observers except observer 2, who spent equal time—that is, around 10 s per image—for virtual and histologically stained slides. These results indicate that there is very similar diagnostic utility between the two image modalities

Table 3.1 Pathology validation study of virtual vs. histochemical staining.

Serial number	Tissue, fixation, type of stain	Pathologist #	Histochemically / Virtually stained	Diagnosis	Time to diagnose
1	Ovary, Frozen section, H&E	1	VS	Adenocarcinoma	30 sec
		2	VS	Borderline serous tumor	15 sec
		3	HS	Mucinous adenocarcinoma	10 min
		4	HS	Adenocarcinoma, ?endometrioid	2 min
2	Ovary, Frozen section, H&E	1	VS	Benign ovary	10 sec
		2	VS	Benign ovary	10 sec
		3	HS	Normal ovary with corpus luteal cyst	15 min
		4	HS	Normal	1 min
3	Salivary Gland, FFPE, H&E	1	VS	Benign salivary glands with mild chronic inflammation	10 sec
		2	VS	Benign parotid tissue	5 sec
		3	HS	Normal salivary gland	1 min
		4	HS	No histopathologic abnormality	1 min
4	Salivary Gland, Frozen section, H&E	1	HS	Pleomorphic adenoma	5 sec
		2	HS	Pleomorphic adenoma	5 sec
		3	VS	Pleomorphic adenoma	3 min
		4	VS	Pleomorphic adenoma	2 sec
5	Salivary Gland, FFPE, H&E	1	HS	Mucoepidermoid carcinoma, low grade	5 sec
		2	HS	Salivary duct carcinoma	5 sec
		3	VS	Mucoepidermoid carcinoma	10 min
		4	VS	Mucoepidermoid Carcinoma	10 sec
6	Breast, FFPE, H&E	1	VS	Invasive ductal carcinoma and DCIS	15 sec
		2	VS	Ductal carcinoma	10 sec
		3	HS	Invasive ductal carcinoma with DCIS	2 min
		4	HS	Invasive carcinoma	1 minute
7	Skin, FFPE, H&E	1	HS	Malignant melanoma	30 sec
		2	HS	melanoma	30 sec
		3	VS	Melanoma	5 min
		4	VS	Melanoma	1 min
8	Prostate, FFPE, H&E	1	HS	Prostatic adenocarcinoma 3+4	1 min
		2	HS	Prostatic adenocarcinoma 4+3	5 sec
		3	VS	Prostatic adenocarcinoma, Gleason pattern 3+4	5 min
		4	VS	HG-PIN with cribiforming vs carcinoma?	5 min
9	Liver, FFPE, Masson's trichrome	1	VS	Benign liver with mild steatosis	10 sec
		2	VS	Benign liver with steatosis	5 sec
		3	HS	Hepatosteatosis, predominantly macrovesicular	3 min
		4	HS	Minimal steatosis, no fibrosis	5 min
10	Liver, FFPE, Masson's trichrome	1	HS	Benign liver with bridging fibrosis	10 sec
		2	HS	Benign liver, bridging fibrosis	5 sec
		3	VS	Moderate cirrhosis	1 min
		4	VS	Mild portal inflammation, focal bridging fibrosis (Stage 2-3)	5 minutes
11		1	VS	Carcinoma	5 sec
		2	VS	Intraductal ca	20 sec

	Salivary Gland, FFPE, H&E	3	HS	Poorly differentiated carcinoma	1 min
		4	HS	Low-grade salivary gland neoplasm	1 minute
12	Salivary Gland, FFPE, H&E	1	HS	Adenocarcinoma	5 sec
		2	HS	Salivary duct carcinoma	5 sec
		3	VS	Salivary duct carcinoma	2 min
		4	VS	Low-grade salivary gland neoplasm	1 minute
13	Thyroid, FFPE, H&E	1	VS	Papillary thyroid carcinoma, tall cell type	10 sec
		2	VS	Papillary thyroid ca, tall cell	20 sec
		3	HS	Papillary thyroid carcinoma, tall cell variant	5 min
		4	HS	PTC	10 sec
14	Thyroid, FFPE, H&E	1	HS	Papillary thyroid carcinoma	5 sec
		2	HS	Medullary ca	5 sec
		3	VS	Papillary thyroid carcinoma, oncocytic variant	7 min
		4	VS	PTC	10 sec
15	Thyroid, FFPE, H&E	1	VS	Papillary thyroid carcinoma	5 sec
		2	VS	Papillary thyroid ca	5 sec
		3	HS	Papillary thyroid carcinoma	1 min
		4	HS	PTC	10 sec

Table 3.2 Blind evaluation of virtual and histological Masson's trichrome staining of liver tissue. for nuclear detail (ND), cytoplasmic detail (CD) and extracellular fibrosis (EF) and overall stain (SQ) score. 4 = perfect, 3 = very good, 2 = acceptable, 1 = unacceptable. The winner (and tied) average scores are bolded.

Tissue #	Pathologist 1				Pathologist 2				Pathologist 3				Average			
	ND	CD	EF	SQ	ND	CD	EF	SQ	ND	CD	EF	SQ	ND	CD	EF	SQ
1 – HS	3	2	1	1	4	4	3	4	1	1	1	3	**2.67**	2.33	1.67	2.67
1 - VS	3	3	3	3	3	3	2	3	2	2	3	3	**2.67**	**2.67**	**2.67**	**3.00**
2 – HS	3	2	4	4	4	4	3	4	1	2	2	2	2.67	**2.67**	3.00	**3.33**
2 - VS	3	3	4	4	4	3	3	3	2	2	3	3	**3.00**	**2.67**	**3.33**	**3.33**
3 – HS	3	3	2	2	3	3	4	3	1	1	1	1	**2.33**	**2.33**	**2.33**	**2.00**
3 - VS	3	2	1	1	3	3	1	4	1	1	1	1	**2.33**	2.00	1.00	**2.00**
4 – HS	3	2	4	4	3	4	4	4	1	2	1	2	2.33	**2.67**	3.00	3.33
4 - VS	3	3	4	4	4	3	4	4	2	2	3	3	**3.00**	**2.67**	**3.67**	**3.67**
5 – HS	3	3	4	4	3	3	2	1	1	3	2	2	2.33	**3.00**	2.67	2.33
5 - VS	3	2	3	3	3	3	4	2	2	1	3	3	**2.67**	2.00	**3.33**	**2.67**
6 – HS	3	2	3	3	4	4	4	3	2	2	2	2	**3.00**	**2.67**	**3.00**	**2.67**
6 - VS	3	3	4	3	4	3	4	3	1	1	1	1	2.67	2.33	**3.00**	2.33
7 – HS	3	3	4	4	3	4	4	3	2	1	2	2	2.67	**2.67**	**3.33**	**3.00**
7 - VS	3	2	3	3	4	4	4	3	2	2	3	3	**3.00**	**2.67**	**3.33**	**3.00**
8 – HS	3	3	4	4	4	4	4	3	1	1	1	1	2.67	**2.67**	3.00	2.67
8 - VS	3	2	4	4	4	3	4	4	2	2	3	2	**3.00**	2.33	**3.67**	**3.33**

Table 3.3 Blind evaluation of virtual and histological Jones staining of kidney tissue sections. Evaluation of nuclear detail (ND), cytoplasmic detail (CD) and overall stain quality (SQ) score. 4 = perfect, 3 = very good, 2 = acceptable, 1 = unacceptable. The winner (and tied) average scores are bolded.

Tissue #	Pathologist 1			Pathologist 2			Pathologist 3			Average		
	ND	CD	SQ	ND	CD	SQ	ND	CD	SQ	ND	CD	SQ
1 – HS	3	3	3	2	2	4	2	2	2	2.33	2.33	3.00
1 - VS	2	3	3	3	3	4	3	3	3	**2.67**	**3.00**	**3.33**
2 – HS	2	4	4	3	3	2	1	1	2	**2.00**	**2.67**	2.67
2 - VS	2	3	4	3	3	3	1	2	3	**2.00**	**2.67**	**3.33**
3 – HS	2	3	3	3	3	2	2	3	4	**2.33**	**3.00**	**3.00**
3 - VS	2	3	3	3	3	3	1	2	3	2.00	2.67	**3.00**
4 – HS	3	3	3	2	2	2	1	2	3	**2.00**	**2.33**	**2.67**
4 - VS	3	3	3	2	2	3	1	2	2	**2.00**	**2.33**	**2.67**
5 – HS	3	3	2	3	3	1	3	3	3	3.00	**3.00**	2.00
5 - VS	3	3	2	4	3	4	3	3	4	**3.33**	**3.00**	**3.33**
6 – HS	2	3	3	3	3	1	2	2	2	**2.33**	**2.67**	2.00
6 - VS	2	2	3	2	2	2	2	2	2	2.00	2.00	**2.33**
7 – HS	3	3	2	3	2	2	3	3	3	3.00	**2.67**	**2.33**
7 - VS	3	3	2	4	3	1	3	2	3	**3.33**	**2.67**	2.00

3.3 Staining standardization.

An interesting by-product of the virtual staining approach can be staining standardization. In other words, the deep network converges to a "common stain" colorization [163] scheme as can be observed in **Figure 3.5**, which compares WSIs of histologically and virtually stained liver tissue sections. As illustrated in **Figure 3.5**, the variation in the histologically stained liver tissue sections is higher than that of the virtually stained tissue images. The colorization of the virtual stain is solely the result of its training (i.e., the gold standard histological staining used during the training phase) and can be further adjusted based on the preferences of pathologists, by retraining the network with a new stain colorization. Such "improved" training can be created from scratch or accelerated through transfer learning [93]. This potential staining standardization using deep learning can remedy the negative effects of human-to-human variations at different stages of the sample preparation [94], create a common ground among different clinical laboratories, enhance the diagnostic workflow for clinicians as well as assist the development of new algorithms such as automatic tissue metastasis detection [155] or grading of different types of cancer, among others.

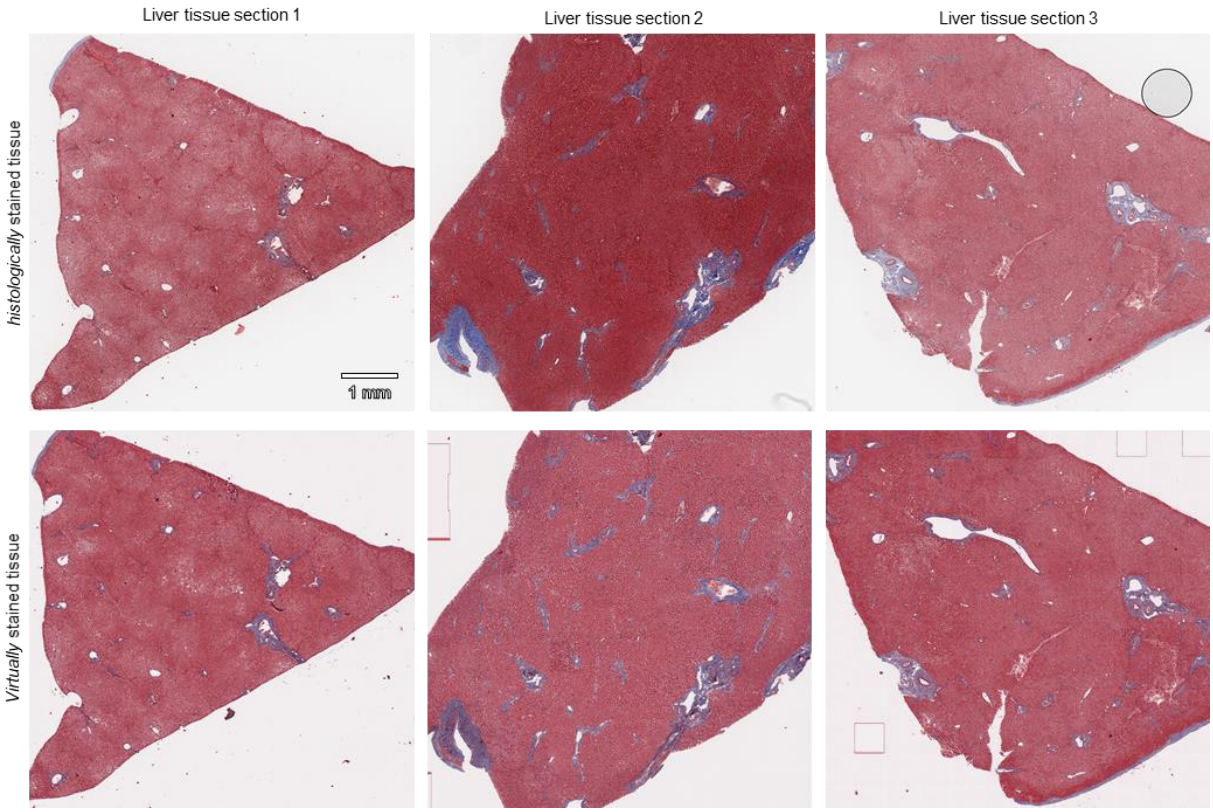

Figure 3.5 Virtual staining reduces staining variability. Staining standardization for whole-slide imaging of liver tissue sections, showing histologically stained and virtually stained images. The virtual staining approach can help to mitigate the staining variability that is part of the histological staining process.

3.4 Transfer learning to other tissue-stain combinations.

Using the concept of transfer learning [93], the training procedure for new tissue and/or stain types can converge much faster, while also reaching an improved performance, i.e., a better local minimum in the training cost/loss function (see the Methods section). This means, a pre-learnt CNN model, from a different tissue-stain combination, can be used to initialize the deep network to statistically learn virtual staining of a new combination. **Figure 3.6** demonstrates the favorable attributes of such an approach: a new deep neural network was trained to virtually stain the auto-fluorescence images of unstained *thyroid* tissue sections, and it was initialized using the weights and biases of another network that was previously trained for H&E virtual staining of the *salivary*

gland. The evolution of the loss metric as a function of the number of iterations used in the training phase clearly demonstrates that the new thyroid deep network rapidly converges to a lower minimum in comparison to the same network architecture which was trained from scratch, using random initialization. **Figure 3.6** also compares the output images of this thyroid network at different stages of its learning process, which further illustrates the impact of transfer learning to rapidly adapt the presented approach to new tissue/stain combinations. The network output images, after the training phase with e.g., $\geq$ 6,000 iterations, reveal that cell nuclei show irregular contours, nuclear grooves, and chromatin pallor, suggestive of papillary thyroid carcinoma; cells also show mild to moderate amounts of eosinophilic granular cytoplasm and the fibrovascular core at the network output image shows increased inflammatory cells including lymphocytes and plasma cells.

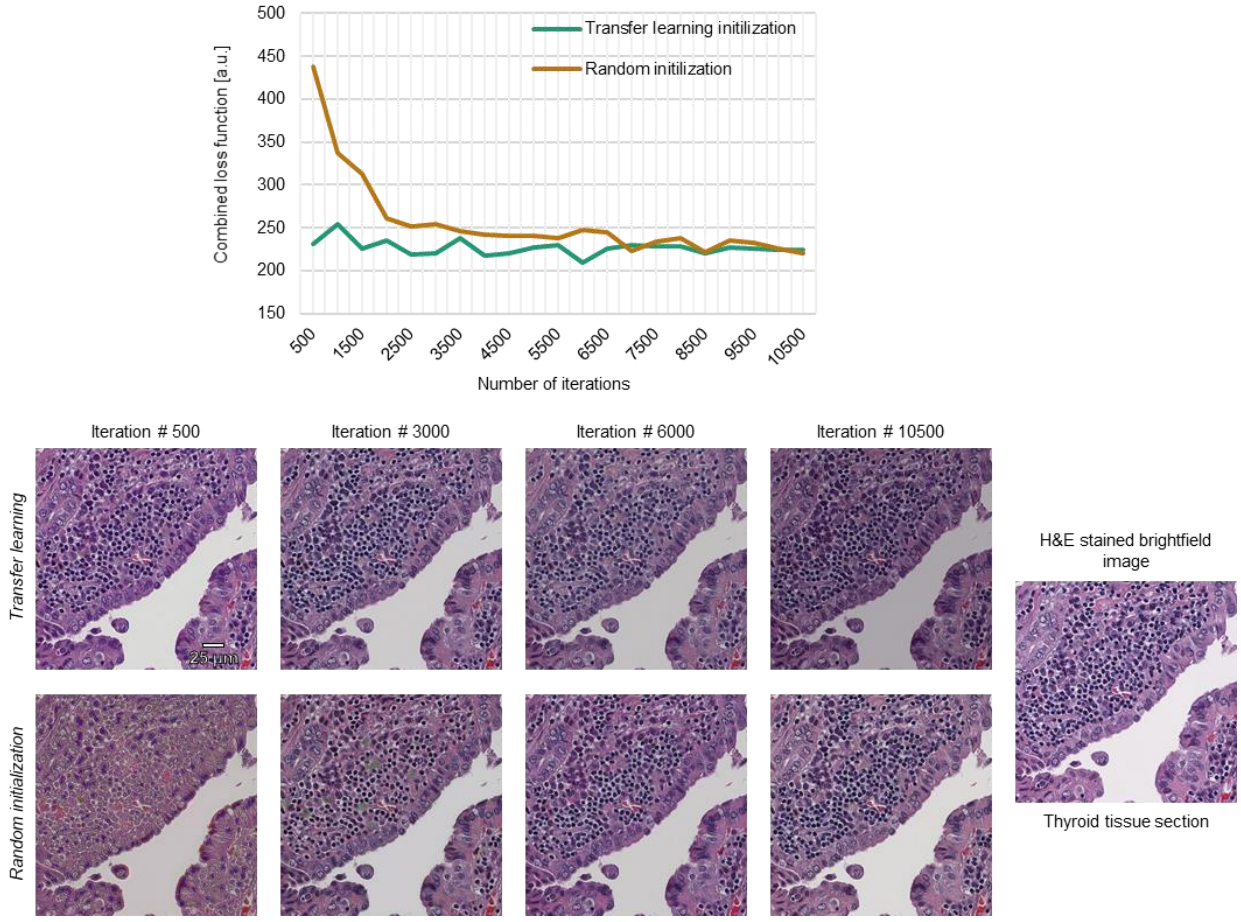

Figure 3.6 Accelerated convergence is achieved using transfer learning. **a**, Plot showing combined loss function against stage of the learning process. A new deep neural network is initialized using the weights and biases learned from the salivary gland tissue sections (see Fig. 3) to achieve virtual staining of thyroid tissue with H&E. Compared to random initialization, transfer learning enables faster convergence, achieving a lower local minimum. **b**, Network output images of a thyroid tissue section at different stages of the learning process, that is, after 500 iterations, 3,000 iterations, 6,000 iterations and 10,500 iterations (left) and a bright-field image of the same thyroid section stained with H&E. The images are compared to each other to better illustrate the impact of the transfer learning method to translate our approach to new tissue–stain combinations.

3.5 Material and methods

Sample preparation

Formalin-fixed paraffin-embedded 2 µm thick tissue sections were deparaffinized using Xylene and mounted on a standard glass slide using Cytoseal^{TM} (Thermo-Fisher Scientific, Waltham, MA USA), followed by placing a coverslip (Fisherfinest, 24x50-1, Fisher Scientific, Pittsburgh, PA USA). Following the initial auto-fluorescence imaging process (using a DAPI excitation and emission filter set) of the unlabeled tissue sample, the slide was then put into Xylene for approximately 48 hours or until the coverslip can be removed without damaging the tissue. Once the coverslip is removed the slide was dipped (approximately 30 dips) in absolute alcohol, 95% alcohol and then washed in D.I. water for ~1 min. This step was followed by the corresponding staining procedures, used for H&E, Masson's Trichrome or Jones stains. This tissue processing path is only used for the training and validation of the approach and is *not needed* after the network has been trained. Different tissue and stain combinations were used to test the virtual staining method: the salivary gland and thyroid tissue sections were stained with H&E, kidney tissue sections were stained with Jones stain, while the liver and lung tissue sections were stained with Masson's trichrome. For the WSI staining efficacy evaluation study, the liver tissue sections were 4 µm thick and the kidney tissue sections were 2 µm thick. In the WSI study, the FFPE tissue sections were not coverslipped during the autofluorescence imaging stage. Following the autofluorescence imaging, the tissue samples were histologically stained as described above (Masson's Trichrome for the liver and Jones for the kidney tissue sections). The unstained frozen samples were prepared by embedding the tissue section in O.C.T. (Tissue Tek, SAKURA FINETEK USA INC) and dipped in 2-Methylbutane with dry ice. The frozen section was then cut to 4 µm sections and was put in a freezer until it was imaged. Following the imaging process, the

tissue section was washed with 70% alcohol, H&E stained and coverslipped. The samples were obtained from the Translational Pathology Core Laboratory (TPCL) and were prepared by the Histology Lab at UCLA. The kidney tissue sections of diabetic and non-diabetic patients were obtained under IRB 18-001029 (UCLA). All the samples were obtained after de-identification of the patient related information and were prepared from existing specimen. Therefore, this work did not interfere with standard practices of care or sample collection procedures.

Data acquisition

The label-free tissue auto-fluorescence images were captured using a conventional fluorescence microscope (IX83, Olympus Corporation, Tokyo, Japan) equipped with a motorized stage, where the image acquisition process was controlled by MetaMorph® microscope automation software (Molecular Devices, LLC). The unstained tissue samples were excited with near UV light and imaged using a DAPI filter cube (OSFI3-DAPI-5060C, excitation wavelength 377 nm / 50 nm bandwidth, emission wavelength 447 nm / 60 nm bandwidth) with a 40×/0.95NA objective lens (Olympus UPLSAPO 40X2/0.95NA, WD0.18) or 20×/0.75NA objective lens (Olympus UPLSAPO 20X/0.75NA, WD0.65). For the melanin inference, the autofluorescence images of the samples using a Cy5 filter cube (CY5-4040C-OFX, excitation wavelength 628 nm / 40 nm bandwidth, emission wavelength 692 nm / 40 nm bandwidth) were additionally acquired with a 10×/0.4NA objective lens (Olympus UPLSAPO10X2). Each auto-fluorescence image was captured with a scientific CMOS sensor (ORCA-flash4.0 v2, Hamamatsu Photonics K.K., Shizuoka Prefecture, Japan) with an exposure time of ~50-500 ms for the DAPI channel and ~3 sec for the Cy5 channel (due to its lower NA). The bright-field images (used for the training and validation) were acquired using a slide scanner microscope (Aperio AT, Leica Biosystems) using a 20×/0.75NA objective (Plan Apo), equipped with a 2× magnification adapter.

Image pre-processing and alignment

Since the deep neural network aims to learn a statistical transformation between an auto-fluorescence image of an unstained tissue and a bright-field image of the same tissue sample after the histological staining, it is of critical importance to accurately match the FOV of the input and target images. An overall scheme describing the global and local image registration process is described in **Figure 3.7**, which was implemented in Matlab (The MathWorks Inc., Natick, MA, USA). The first step in this process is to find candidate features for matching unstained auto-fluorescence images and stained bright-field images. For this, each auto-fluorescence image (2048×2048 pixels) is down-sampled to match the effective pixel size of the bright-field microscope images. This results in a 1351×1351-pixel unstained auto-fluorescent tissue image, which is contrast enhanced by saturating the bottom 1% and the top 1% of all the pixel values, and contrast reversed to better represent the color map of the grayscale converted whole slide image (**Figure 3.7**). Then, a normalized correlation score matrix is calculated by correlating each one of the 1351×1351-pixel patches with the corresponding patch of the same size, extracted from the whole slide gray-scale image. The entry in this matrix with the highest score represents the most likely matched FOV between the two imaging modalities. Using this information (which defines a pair of coordinates), the matched FOV of the original whole slide bright-field image was cropped to create target images. Following this FOV matching procedure, the auto-fluorescence and bright-field microscope images are coarsely matched. However, they are still not accurately registered at the individual pixel-level, due to the slight mismatch in the sample placement at the two different microscopic imaging experiments (auto-fluorescence, followed by bright-field), which randomly causes a slight rotation angle (e.g., ~1-2 degrees) between the input and target images of the same sample.

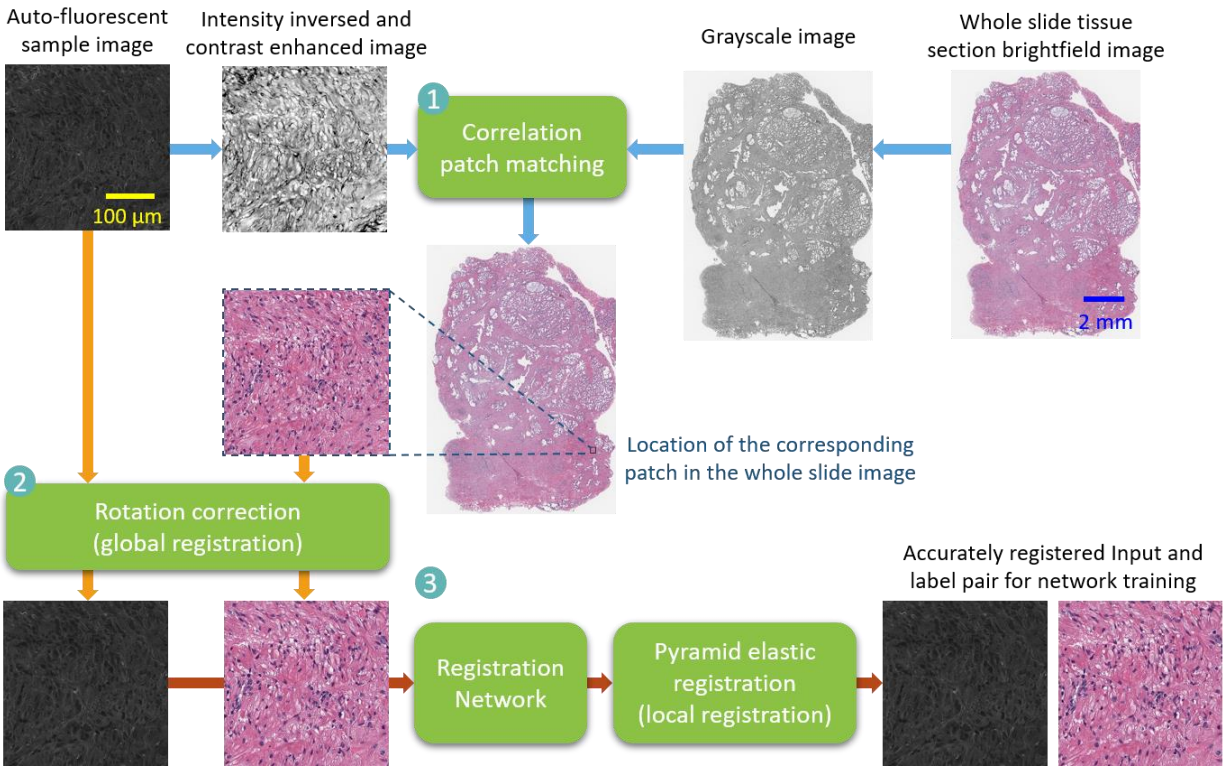

Figure 3.7. Auto-fluorescence and bright-field image registration. The field-of-view matching and registration process of the auto-fluorescence images of unstained tissue samples with respect to the bright-field images of the same samples, after the histological staining process.

The second part of the input-target matching process involves a global registration step [164], which corrects for this slight rotation angle between the auto-fluorescence and bright-field images. This is done by extracting feature vectors (descriptors) and their corresponding locations from the image pairs, and matching the features by using the extracted descriptors [83]. Then, a transformation matrix corresponding to the matched pairs is found using the M-estimator Sample Consensus (MSAC) algorithm [165], which is a variant of the Random Sample Consensus (RANSAC) algorithm [166]. Finally, the angle-corrected image is obtained by applying this transformation matrix to the original bright-field microscope image patch. Following the application of this rotation, the images are further cropped by 100 pixels (50 pixels on each side)

to accommodate for undefined pixel values at the image borders, due to the rotation angle correction.

Next, a neural network is used to learn the transformation between the roughly matched images. This network uses the same structure as the network described in **Figure 3.2**. A low number of iterations is used so that the network only learns color mapping, and not any spatial transformations between the input and label images. The auto-fluorescence images are passed through this network and used to perform local feature registration, using an elastic image registration algorithm. This algorithm matches the local features of both sets of images (auto-fluorescence vs. bright-field), by hierarchically matching the corresponding blocks, from large to small (**Figure 3.7**). The calculated transformation map from this step is finally applied to each bright-field image patch [167].

At the end of these registration steps, the auto-fluorescence image patches and their corresponding bright-field tissue image patches are accurately matched to each other and can be used as input and label pairs for the deep neural network training phase, allowing the network to *solely* focus on and learn the problem of virtual histological staining.

For the 20× objective lens images (that were used for **Table 3.2** and **Table 3.3**) a similar process was used. Instead of down-sampling the auto-fluorescence images, the bright-field microscope images were down-sampled to 75.85% of their original size so that they match with the lower magnification images. Furthermore, to create whole slide images using these 20× images, additional shading correction and normalization techniques were applied. Before being fed into the network, each field-of-view was normalized by subtracting the mean value across the entire slide and dividing it by the standard deviation between pixel values. This normalizes the network

input both within each slide as well as between slides. Finally, shading correction was applied to each image to account for the lower relative intensity measured at the edges of each field-of-view.

Deep neural network architecture, training, and validation

In this work, a GAN [157] architecture was used to learn the transformation from a label-free unstained auto-fluorescence input image to the corresponding bright-field image of the histologically stained sample. A standard convolutional neural network-based training learns to minimize a loss/cost function between the network's output and the target label. Thus, the choice of this loss function is a critical component of the deep network design. For instance, simply choosing an ℓ_2-norm penalty as a cost function will tend to generate blurry results [168,169], as the network averages a weighted probability of all the plausible results; therefore, additional regularization terms [79,170] are generally needed to guide the network to preserve the desired sharp sample features at the network's output. GANs avoid this problem by learning a criterion that aims to accurately classify if the deep network's output image is real or fake (i.e., correct in its virtual staining or wrong). This makes the output images that are inconsistent with the desired labels not to be tolerated, which makes the loss function to be *adaptive* to the data and the desired task at hand. To achieve this goal, the GAN training procedure involves training of two different networks, as shown in **Figure 3.2**: (*i*) a *generator* network, which in this case aims to learn the statistical transformation between the unstained auto-fluorescence input images and the corresponding bright-field images of the same samples, after the histological staining process; and (*ii*) a *discriminator* network that learns how to discriminate between a true bright-field image of a stained tissue section and the generator network's output image. Ultimately, the desired result of this training process is a generator, which transforms an unstained auto-fluorescence input image

into an image which will be *indistinguishable* from the stained bright-field image of the same sample. For this task, the loss functions of the generator and discriminator were defined as such:

$$\begin{aligned} \ell_{\text{generator}} &= \text{MSE}\left\{z_{\text{label}}, z_{\text{output}}\right\} + \lambda \times \text{TV}\left\{z_{\text{output}}\right\} + \alpha \times (1 - D(z_{\text{output}}))^2 \\ \ell_{\text{discrimnator}} &= D(z_{\text{output}})^2 + (1 - D(z_{\text{label}}))^2 \end{aligned} \tag{3.1}$$

where D refers to the discriminator network output, z_{label}denotes the bright-field image of the histologically stained tissue, z_{output} denotes the output of the generator network. The generator loss function balances the pixel-wise mean squared error (MSE) of the generator network output image with respect to its label, the total variation (TV) operator of the output image, and the discriminator network prediction of the output image, using the regularization parameters (λ, α) that are empirically set to different values, which accommodate for ~2% and ~20% of the pixel-wise MSE loss and the combined generator loss ($\ell_{\text{generator}}$), respectively. The TV operator of an image z is defined as:

$$\text{TV}(z) = \sum_p \sum_q \sqrt{\left(z_{p+1,q} - z_{p,q}\right)^2 + \left(z_{p,q+1} - z_{p,q}\right)^2} \tag{3.2}$$

where p, q are pixel indices. Based on Eq. (1), the discriminator attempts to minimize the output loss, while maximizing the probability of correctly classifying the real label (i.e., the bright-field image of the histologically stained tissue). Ideally, the discriminator network would aim to achieve $D(z_{\text{label}}) = 1$ and $D(z_{\text{output}}) = 0$, but if the generator is successfully trained by the GAN, $D(z_{\text{output}})$ will ideally converge to 0.5.

The generator deep neural network architecture follows the design of U-net [171], and is detailed in **Figure 3.2**. The U-net architecture is well suited for this application because it is capable of learning features at different scales without increasing the depth of the network. Each

level of the U-net downsamples the input and learns the features that act on a larger scale than that of the previous layer. This allows the network to infer small features within each cell as well as the overall structure of the tissue samples. An input image is processed by the network in a multi-scale fashion, using down-sampling and up-sampling paths, helping the network to learn the virtual staining task at various different scales. The down-sampling path consists of four individual steps, with each step containing one residual block [172], each of which maps a feature map x_k into feature map x_{k+1}:

$$x_{k+1} = x_k + \mathrm{LReLU}\left[\mathrm{CONV}_{k3}\left\{\mathrm{LReLU}\left[\mathrm{CONV}_{k2}\left\{\mathrm{LReLU}\left[\mathrm{CONV}_{k1}\left\{x_k\right\}\right]\right\}\right]\right\}\right] \tag{3.3}$$

where CONV{.} is the convolution operator (which includes the bias terms), *k1*, *k2*, and *k3* denote the serial number of the convolution layers, and LReLU[.] is the non-linear activation function (i.e., a Leaky Rectified Linear Unit) that was used throughout the entire network, defined as:

$$\mathrm{LReLU}(x) = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \tag{3.4}$$

When training the networks for whole slide images, an additional batch normalization layer was added before each LReLU activation to allow for faster training and improve its stability. This addition particularly improves sections of the tissue, where the contrast in the auto-fluorescence images is particularly low. The number of the input channels for each level in the down-sampling path was set to: 1, 64, 128, 256, while the number of the output channels in the down-sampling path was set to: 64, 128, 256, 512. To avoid the dimension mismatch for each block [79], feature map x_k was zero-padded to match the number of the channels in x_{k+1} The connection between each down-sampling level is a 2×2 average pooling layer with a stride of 2 pixels that down-samples the feature maps by a factor of 4 (2-fold for in each direction). Following the output of

the fourth down-sampling block, another convolutional layer maintains the number of the feature maps as 512, before connecting it to the up-sampling path. The up-sampling path consists of four, symmetric, up-sampling steps, with each step containing one convolutional block. The convolutional block operation, which maps feature map y_k into feature map y_{k+1}, is given by:

$$y_{k+1} = \mathrm{LReLU}\left[\mathrm{CONV}_{k6}\left\{\mathrm{LReLU}\left[\mathrm{CONV}_{k5}\left\{\mathrm{LReLU}\left[\mathrm{CONV}_{k4}\left\{\mathrm{CONCAT}\left(x_{k+1}, \mathrm{US}\{y_k\}\right)\right\}\right]\right\}\right]\right\}\right] \tag{3.5}$$

where CONCAT(.) is the concatenation between two feature maps which merges the number of channels, US{.} is the up-sampling operator, and *k*4, *k*5, and *k*6, denote the serial number of the convolution layers. Similar to the down-sampling path, batch normalization was added for the whole slide image training phase. The number of the input channels for each level in the up-sampling path was set to 1024, 512, 256, 128 and the number of the output channels for each level in the up-sampling path was set to 256, 128, 64, 32, respectively. The last layer is a convolutional layer mapping 32 channels into 3 channels, represented by the YCbCr color map [173]. Both the generator and the discriminator networks were trained with a patch size of 256×256 pixels.

The discriminator network, summarized in **Figure 3.2**, receives 3 input channels, corresponding to the YCbCr color space of an input image. This input is then transformed into a 64-channel representation using a convolutional layer, which is followed by 5 blocks of the following operator:

$$z_{k+1} = \mathrm{LReLU}\left[\mathrm{CONV}_{k2}\left\{\mathrm{LReLU}\left[\mathrm{CONV}_{k1}\left\{z_k\right\}\right]\right\}\right] \tag{3.6}$$

where *k*1, *k*2, denote the serial number of the convolutional layer. The number of channels for each layer was 3, 64, 64, 128, 128, 256, 256, 512, 512, 1024, 1024, 2048. The next layer was an average

pooling layer with a filter size that is equal to the patch size (256×256), which results in a vector with 2048 entries. The output of this average pooling layer is then fed into two fully connected layers with the following structure:

$$z_{k+1} = \mathrm{FC}\left[\mathrm{LReLU}\left[\mathrm{FC}\left\{z_k\right\}\right]\right] \tag{3.7}$$

where FC represents the fully connected layer, with learnable weights and biases. The first fully connected layer outputs a vector with 2048 entries, while the second one outputs a scalar value. This scalar value is used as an input to a sigmoid activation function $D(z) = 1/(1 + \exp(-z))$ which calculates the probability (between 0 and 1) of the discriminator network input to be real/genuine or fake, i.e., ideally $D(z_{\mathrm{label}}) = 1$.

The convolution kernels throughout the GAN were set to be 3×3. These kernels were randomly initialized by using a truncated normal distribution [71] with a standard deviation of 0.05 and a mean of 0; all the network biases were initialized as 0. The learnable parameters are updated through the training stage of the deep network using an adaptive moment estimation (Adam) optimizer [174] with learning rate 1×10^{-4} for the generator network and 1×10^{-5} for the discriminator network. Also, for each iteration of the discriminator, there were 4 iterations of the generator network, to avoid training stagnation following a potential over-fit of the discriminator network to the labels. A batch size of 10 has been used in the training.

Once all the fields-of-view have passed through the network, the whole slide images are stitched together using the Fiji [175] Grid/Collection stitching plugin [176]. This plugin calculates the exact overlap between each tile and linearly blends them into a single large image. Overall, the inference and stitching took ~5 minutes and 30 seconds, respectively, per cm^2 and can be substantially improved using hardware and software advancements. Before being shown to the

pathologists, sections which are out of focus or have major aberrations (due to e.g., dust particles) in either the auto-fluorescence or bright-field images are cropped out. Finally, the images were exported to the Zoomify [177] format (designed to enable viewing of large images using a standard web browser) and uploaded to the GIGAmacro website [178] for easy access and viewing by the pathologists.

Implementation details

The other implementation details, including the number of trained patches, the number of epochs and the training times are shown in **Table 3.4** Training details for different tissue/stain combinations.. The virtual staining network was implemented using Python version 3.5.0. The GAN was implemented using TensorFlow framework version 1.4.0. Other python libraries used were os, time, tqdm, the Python Imaging Library (PIL), SciPy, glob, ops, sys, and numpy. The software was implemented on a desktop computer with a Core i7-7700K CPU @ 4.2GHz (Intel) and 64GB of RAM, running a Windows 10 operating system (Microsoft). The network training and testing were performed using dual GeForce GTX 1080Ti GPUs (NVidia).

Table 3.4 Training details for different tissue/stain combinations.

Virtual staining network	# of training patches	# of epochs	Training time (hours)
Salivary gland (H&E)	2768	26	13.046
Thyroid (H&E)	8336	8	12.445
Thyroid (H&E, transfer learning)	8336	4	7.107
Liver (Masson's Trichrome)	3840	26	18.384
Lung (Masson's Trichrome)	9162	10	16.602
Kidney (Jones stain)	4905	8	7.16
Liver (Masson's Trichrome, WSI)	211475	3	39.64
Kidney (Jones stain, WSI)	59344	14	57.05

Ovary 1	4738	84	37.21
Ovary 2	11123	14	37.41
Salivary Gland - 1	4417	65	24.61
Salivary Gland – 2	2652	90	23.9
Salivary Gland – 3	13262	24	30.58
Breast	67188	4	24.85
Skin	2566	124	27.02
Skin (DAPI+CY5)	2566	124	29.62
Prostate	677	472	30.27

3.6 Discussion

The ability to virtually stain label-free tissue sections was demonstrated using a supervised deep learning technique that uses a single auto-fluorescence image of the sample as input, captured by a standard fluorescence microscope and filter set. This statistical learning-based method has the potential to restructure the clinical workflow in histopathology and can benefit from various imaging modalities such as fluorescence microscopy, non-linear microscopy, holographic microscopy and optical coherence tomography [179], among others, to potentially provide a digital alternative to the standard practice of histological staining of tissue samples. In this work, this method was demonstrated using fixed unstained tissue samples to provide a meaningful comparison to histologically stained tissue samples, which is essential to train the neural network as well as to blindly test the performance of the network output against the clinically approved method. However, the presented deep learning-based approach is broadly applicable to un-sectioned, fresh tissue samples without the use of any labels or stains. Following its training, the deep network can be used to virtually stain the images of label-free fresh tissue samples, acquired using e.g., UV or deep UV excitation or even nonlinear microscopy modalities. Especially, Raman

microscopy can provide very rich label-free biochemical signatures that can further enhance the effectiveness of the virtual staining that the neural network learns.

The proposed method can be combined with other excitation wavelengths and/or imaging modalities in order to enhance its inference performance for different tissue constituents. For example, I attempted to detect melanin on a skin tissue section using virtual H&E staining. However, melanin was not clearly identified in the output of the network, as it presents a weak auto-fluorescent signal at DAPI excitation/emission wavelengths [180] measured in this system. One potential method to increase the autofluorescence of melanin is to image the samples while they are in an oxidizing solution [181]. As a more practical alternative, here an additional autofluorescence channel originating from e.g., Cy5 filter (excitation 628 nm/emission 692 nm) was used such that the melanin signal can be enhanced and accurately inferred in the virtual staining framework. By training the network using both the DAPI and Cy5 autofluorescence channels, the deep network was able to successfully determine where melanin occurs in the sample, as illustrated in **Figure 3.8**. In contrast, when only the DAPI channel was used (**Figure 3.8**a), the network was unable to determine the areas that contain melanin. Stated differently, the additional autofluorescence information from the Cy5 channel was used by the network to distinguish melanin from the background tissue. It should also be noted here that the results that are shown in **Figure 3.8** were acquired using a lower resolution objective lens (10×/0.45NA) for the Cy5 channel, to supplement the high-resolution DAPI scan (20×/0.75NA), as I hypothesized that most necessary information is found in the high-resolution DAPI scan and the additional information (for example, the melanin presence) can be encoded with the lower resolution scan. I believe that other label-free imaging techniques and/or fluorescence channels can be combined to further

enhance the inference of different tissue constituents using the presented deep learning-based approach, which are left as future research.

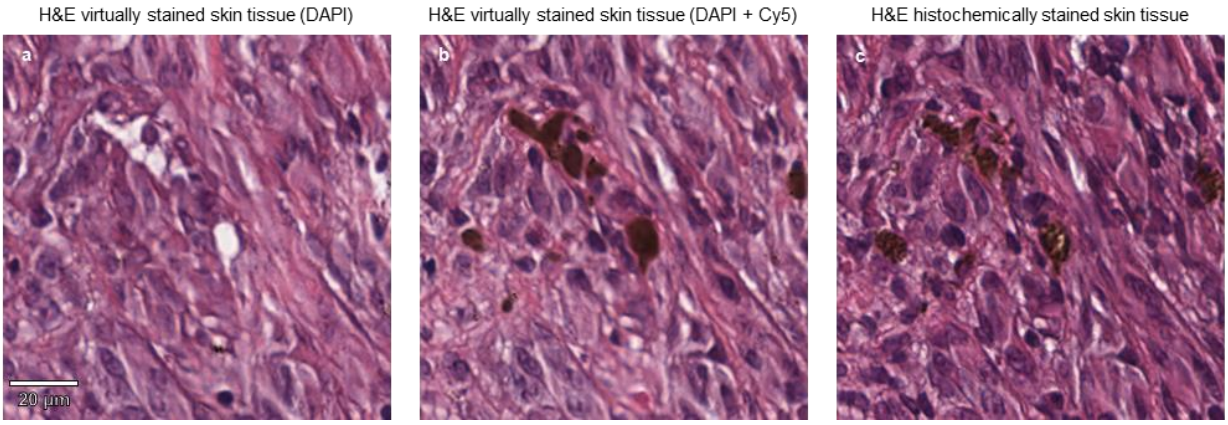

Figure 3.8 Melanin inference using multiple autofluorescence channels. **a**, Virtually stained skin tissue sample, using the DAPI channel only. **b**,**c**, The same tissue sample, virtually stained using both the DAPI and Cy5 channels (**b**), clearly revealing the melanin (dark-brown) features that are shown in the corresponding histologically stained image (**c**).

An important part of the training process involves matching the auto-fluorescence images of label-free tissue samples and their corresponding bright-field images after the histological staining process. One should note that during the staining process and related steps, some tissue constitutes can be lost or deformed in a way that will mislead the loss/cost function in the training phase. This, however, is only a training and validation related challenge and does *not* pose any limitations on the practice of a well-trained neural network for virtual staining of label-free tissue samples. To ensure the quality of the training and validation phases and minimize the impact of this challenge on the network's performance, a multi-stage registration process was applied, from global to local registration, where one of the steps involves training a deep network for the task of enabling high accuracy local registration. At the end of this initial registration, a threshold for an acceptable correlation value between the two sets of images (i.e., before and after the histological staining

process) was set to eliminate the non-matching image pairs from the training/validation set to make sure that the network learns the real signal, not the perturbations to the tissue morphology due to the histological staining process. This threshold is tuned for each specific tissue, depending on the correlation between the input and label images. For example, if there is a large misalignment between the input and the target images of the registration network, the correlation will be low due to this misalignment. Additionally, when the registration of the images that are used as input into the registration network is poor, the network will produce low quality images. This in turn also results a low correlation value (see the Methods section for further details regarding the registration network). This threshold is tuned manually by inspecting images to determine whether improperly co-registered images make it past the threshold and by making sure that the number of images allowed through is sufficient to train the network. As a small number of poorly co-registered images can reduce the quality of the network, a significant safety margin is used wherever possible to eliminate any images that have the possibility of being poorly co-registered.

A methodology was described above to mitigate some of the training challenges due to random loss of some tissue features after the histological staining process. In fact, this highlights another motivation to skip the laborious and costly procedures that are involved in histological staining as it will be easier to preserve the local tissue histology in a label-free method, without the need for an expert to handle some of the delicate procedures of the staining process, which sometimes also requires observing the tissue under a microscope.

The training phase of the deep neural network takes a considerable amount of time (e.g., ~13 hours for the salivary gland network) using a desktop PC; however, this entire process can be significantly accelerated by using dedicated hardware, based on GPUs. Furthermore, as already emphasized in **Figure 3.6**, transfer learning provides a warm start to the training phase of a new

tissue/stain combination, making the entire process significantly faster. Unlike other color reconstruction or virtual staining approaches [156], once the deep network has been trained, the virtual staining of a new sample is performed in a single, non-iterative manner, which does not require a trial-and-error approach or any parameter tuning to achieve the optimal result. Based on its feed-forward and non-iterative architecture, the deep neural network rapidly outputs a virtually stained image in e.g., 1.9 sec/mm^2 using a dual-GPU desktop computer, for unstained tissue slides scanned using a 20× objective lens. With further GPU-based acceleration and machine learning optimized processors, this approach has the potential to achieve real-time performance, which might especially be useful in the operating room or for *in vivo* imaging applications.

The virtual staining procedure that is implemented in this work is based on training a separate CNN for each tissue/stain combination. If one feeds a CNN with the auto-fluorescence images of a different tissue/stain combination, it may not perform as desired. This, however, is *not* a limitation because for histology applications, the tissue and stain type are pre-determined for each sample of interest, and therefore, a specific CNN selection for creating a virtually stained image from an auto-fluorescence image of the unlabeled sample does not require an additional information or resource. A more general CNN model can be learnt for multiple tissue/stain combinations by e.g., increasing the number of trained parameters in the model [12], at the cost of a possible increase in the training and inference times. Using a similar strategy, another avenue to explore in future work is the potential of the presented framework to perform multiple virtual stains on the same unlabeled tissue type.

It is important to note that, like in any other imaging method that is based on automatic sample scanning, parts of the sample field-of-view can be compromised due to artifacts in the sample preparation process, such as dust or other particles that lay on top of the sample, in addition to

tissue folding and cracks, among other artifacts. Further development of auto-focusing algorithms that can learn to reject such artifacts during the imaging stage could minimize their occurrence in the final image.

As for the next steps, a wide-scale randomized clinical study would be needed to validate the diagnostic accuracy of the network output images, against the clinical gold standard, which will be important to better understand potential biases in the output images of the network. A significant advantage of the presented framework is that it is quite flexible: it can accommodate feedback to statistically mend its performance if a diagnostic failure is detected through a clinical comparison, by accordingly penalizing such failures as they are caught. This iterative training and transfer learning cycle, based on clinical evaluations of the performance of the network output, will help us optimize the robustness and clinical impact of the presented approach. In this sense, the process bears resemblance to the design phase of a histological stain, where through trial and error the stain is optimized to provide desired contrast to specific histological features.

I would like to also point to another exciting opportunity created by this framework for micro-guiding molecular analysis at the unstained tissue level, by locally identifying regions of interest based on virtual staining, and using this information to guide subsequent analysis of the tissue for e.g., micro-immunohistochemistry or sequencing [161]. This type of virtual micro-guidance on an unlabeled tissue sample can facilitate high-throughput identification of sub-types of diseases, also helping the development of customized therapies for patients [182].

Finally, I would like to note that while the presented virtual staining method was demonstrated for a contrast mechanism that originates from tissue autofluorescence with a single excitation band, other contrast generating methods to virtually stain label-free tissue samples should also be explored, including e.g., multiple excitation and emission wavelengths, as well as other imaging

modalities such as polarization imaging, quantitative phase microscopy, optical coherence tomography, and perhaps combinations of these modalities.

Chapter 4 Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning

4.1 Introduction

The rapid and accurate identification of live microorganisms is of great importance for a wide range of applications [183–190], including drug discovery screening assays [183–185], clinical diagnoses [186], microbiome studies [187,188], and food and water safety [189,190]. Waterborne diseases affect more than 2 billion people worldwide [191], causing a substantial economic burden; for example, the treatment of waterborne diseases costs more than $2 billion annually in the United States (US) alone, with 90 million cases recorded per year [192].

Among waterborne pathogen-related problems, one of the most common public health concerns is the presence of total coliform bacteria and *Escherichia coli* (*E. coli*) in drinking water, which indicates fecal contamination. Analytical methods used to detect *E. coli* and total coliforms are based on culturing the obtained samples on solid agar plates (e.g., the US Environmental Protection Agency (EPA) 1103.1 and EPA 1604 methods) or in liquid media (e.g., Colilert test), followed by visual recognition and counting by an expert, as described in the EPA guidelines [193–195]. While the use of liquid growth media for the detection of fecal coliform bacteria provides high sensitivity and specificity, it requires at least 18 h for the final read-out. The use of solid agar plates is a relatively more cost-effective method and provides flexibility for the volume of the sample to be analyzed, which can vary from 100 mL to several liters by using a membrane filtration technique to enhance the sensitivity. However, this traditional culture-based detection method requires the colonies to grow to a certain macroscopic size for visibility, which often takes 24–48 h in the case of bacterial samples. Alternatively, molecular detection methods [196,197] based on, e.g., the amplification of nucleic acids, can reduce the assay time to

a few hours, but they generally lack the sensitivity for detecting bacteria at very low concentrations, e.g., 1 colony forming unit (CFU) per 100-1000 mL, and are *not* capable of differentiating between live and dead microorganisms. [198] Furthermore, there is *no* EPA-approved nucleic acid-based analytical method [199] for detecting coliforms in water samples.

Overall, there is a strong and urgent need for an automated method that can achieve rapid and high-throughput colony detection with high sensitivity (routinely achieving, e.g., 1 CFU per 100-1000 mL in less than 12 h) to provide a powerful alternative to the currently available EPA-approved gold-standard analytical methods that (1) are slow, take ~24–48 h and (2) require experts to read and quantify samples. To address this important need, various other approaches [200–202] have been investigated for the detection of total coliform bacteria and *E. coli* in water samples, including solid phase cytometry [203], droplet-based micro-optical lens array measurements [204], fluorimetry [205], luminometry [206], and fluorescence microscopy [207]. Despite the fact that these methods provide high sensitivity and some time savings, they cannot handle large sample sizes (e.g., ≥100 mL) or cannot perform the automated classification of bacterial colonies.

To provide a highly sensitive and high-throughput system for the early detection and classification of live microorganisms and colony growth, a time-lapse coherent imaging platform that uses two different deep neural networks (DNNs) for its operation is presented. The first DNN is used to detect bacterial growth as early as possible, and the second DNN is used to classify the type of growing bacteria based on the spatiotemporal features obtained from the coherent images of an incubated agar plate (**Figure 4.1**). In this live bacteria detection system, which is integrated with an incubator, lens-free holographic images of the agar plate sample are captured by a monochromatic complementary metal–oxide–semiconductor (CMOS) image sensor that is mounted on a translational stage. The system rapidly scans the entire area of two separate agar

plates (~ 56.52 cm^2) every 30 min and utilizes these time-resolved holographic images for the accurate detection, classification, and counting of the growing colonies as early as possible (**Figure 4.2**a). This unique system enables high-throughput periodic monitoring of an incubated sample by scanning a 60-mm-diameter agar plate in 87 s with an image resolution of <4 μm; it continuously calculates differential images of the sample of interest for the early and accurate detection of bacterial growth. The spatiotemporal features of each nonstatic object on the plate are continuously analysed using deep learning to yield the count of bacterial growth and to automatically identify the type(s) of bacteria growing on the different parts of the agar plate.

The efficacy of this platform was demonstrated by performing the early detection and classification of three types of bacteria, i.e., *E. coli*, *Klebsiella aerogenes (K. aerogenes)*, and *Klebsiella pneumoniae (K. pneumoniae)*, and achieved a limit of detection (LOD) of ~1 CFU/L in ≤ 9 h of the total test time. Moreover, detection time savings of more than 12 h was achieved compared to the gold-standard EPA methods [208], which usually require at least 24 h to obtain a result. The growth statistics of these three different species was also quantified and provided a detailed growth analysis of each type of bacteria over time. The detection and classification neural network models were built, trained, and validated with ~16,000 individual colonies resulting from 71 independent experiments and were blindly tested with 965 individual colonies collected from 15 independent experiments that were never used in the training phase. In the blind testing, the trained models demonstrated an 80% detection sensitivity within 6–9 h, a 90% detection sensitivity within 7–10 h, and a > 95% detection sensitivity within 12 h, while maintaining ~99.2-100% precision at any time point after 7 h, also achieving correct identification of 80% of all three the species within 7.6–12 h. In terms of the species-specific accuracy of the classification network, within 12 h of incubation, it achieved ~97.2%, ~84.0%, and ~98.5% classification accuracy for *E.*

coli, *K. aerogenes*, and *K. pneumoniae*, respectively. These results confirm the transformative potential of this platform, which not only enables the highly sensitive, rapid and cost-effective detection of live bacteria (with a cost of $0.6 per test, including a culture plate) but also provides a powerful and versatile tool for microbiology research.

This system was demonstrated by monitoring bacterial colony growth within 60-mm-diameter agar plates and quantitatively analysed the capabilities of the platform for early detection of the bacterial growth and classification of bacterial species. To demonstrate its proof-of-concept, I aimed to automatically detect, classify, and count *E. coli* and coliform bacteria in water samples using the deep learning-based platform. Throughout the training and blind testing experiments, I used water suspensions spiked with coliform bacteria, including *E. coli, K. aerogenes*, and *K. pneumoniae*, and chlorine-stressed *E. coli*. A chromogenic agar medium designed for the specific detection and counting of *E. coli* and other coliform bacteria in food and water samples was used as a culture medium for specificity (see the Methods section for details). This chromogenic medium results in a blue colour for *E. coli* colonies and a mauve colour for the colonies of other coliform bacteria (e.g., *K. aerogenes* and *K. pneumoniae)*. Additionally, the medium inhibits the growth of different bacteria (e.g., *Bacillus subtilis*) or yields colourless colonies in the presence of other bacteria in the sample [209].

Following the sample preparation method illustrated in **Figure 4.2**a, the sample is placed inside the lens-free imaging system with the agar surface facing the image sensor. After an initialization step, the platform automatically captures time-lapsed holographic images of two separate Petri dishes (covering a total sample area of $28.26 \times 2 = 56.52$ cm^2) every 30 min over a duration of 24 h starting from the incubation time; these individual holograms are digitally stitched together and rapidly reconstructed to reveal the bacterial growth patterns on the agar surface (see

the Methods section). The reconstructed images of the sample captured at different time points are computationally processed using a differential image analysis method to automatically detect and classify bacterial growth and colonies using two different trained DNNs (**Figure 4.3**), which will be detailed next. Part of this chapter has been previously published in :

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

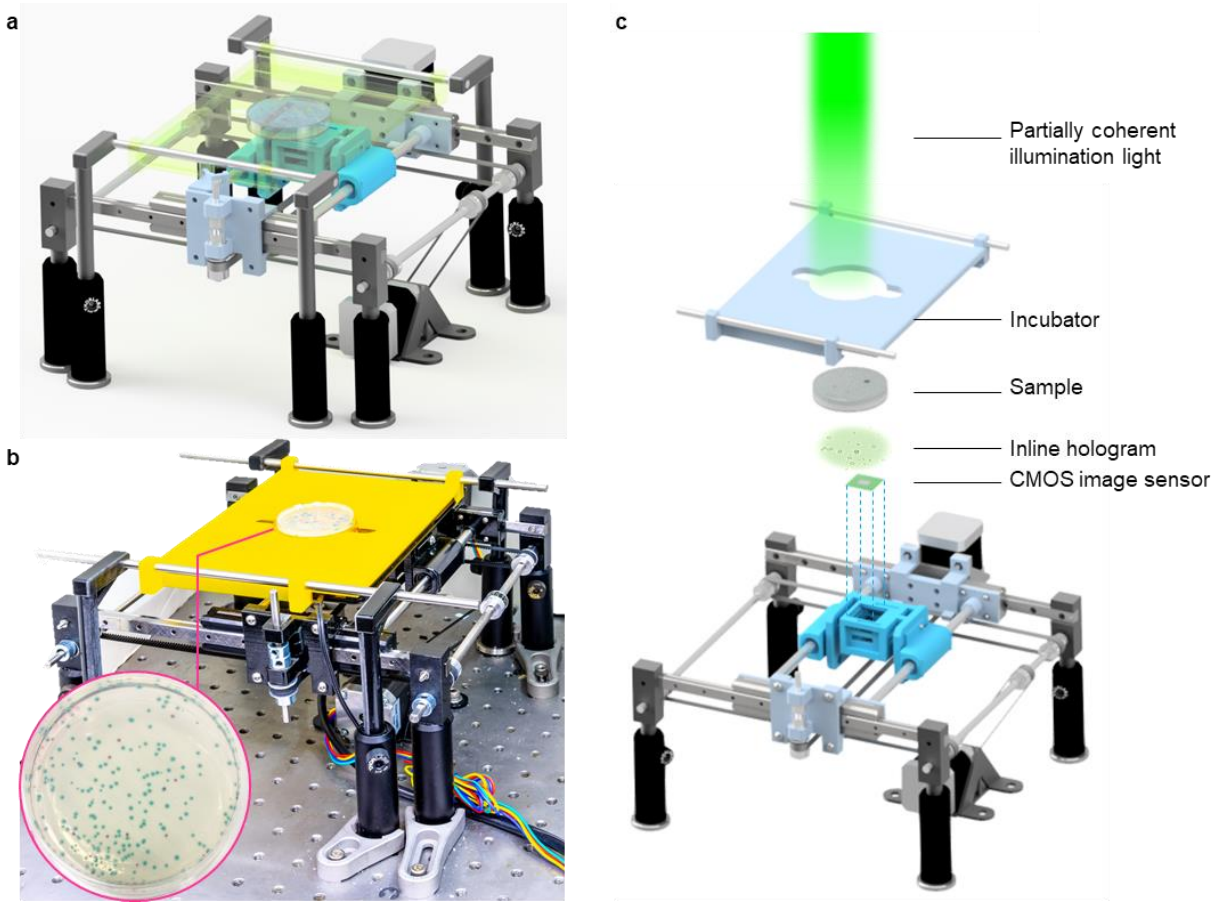

Figure 4.1 High-throughput bacterial colony growth detection and classification system. (a) Schematic of the device. (b) Photograph of the lens-free imaging system. (c) Detailed illustration of various components of the system.

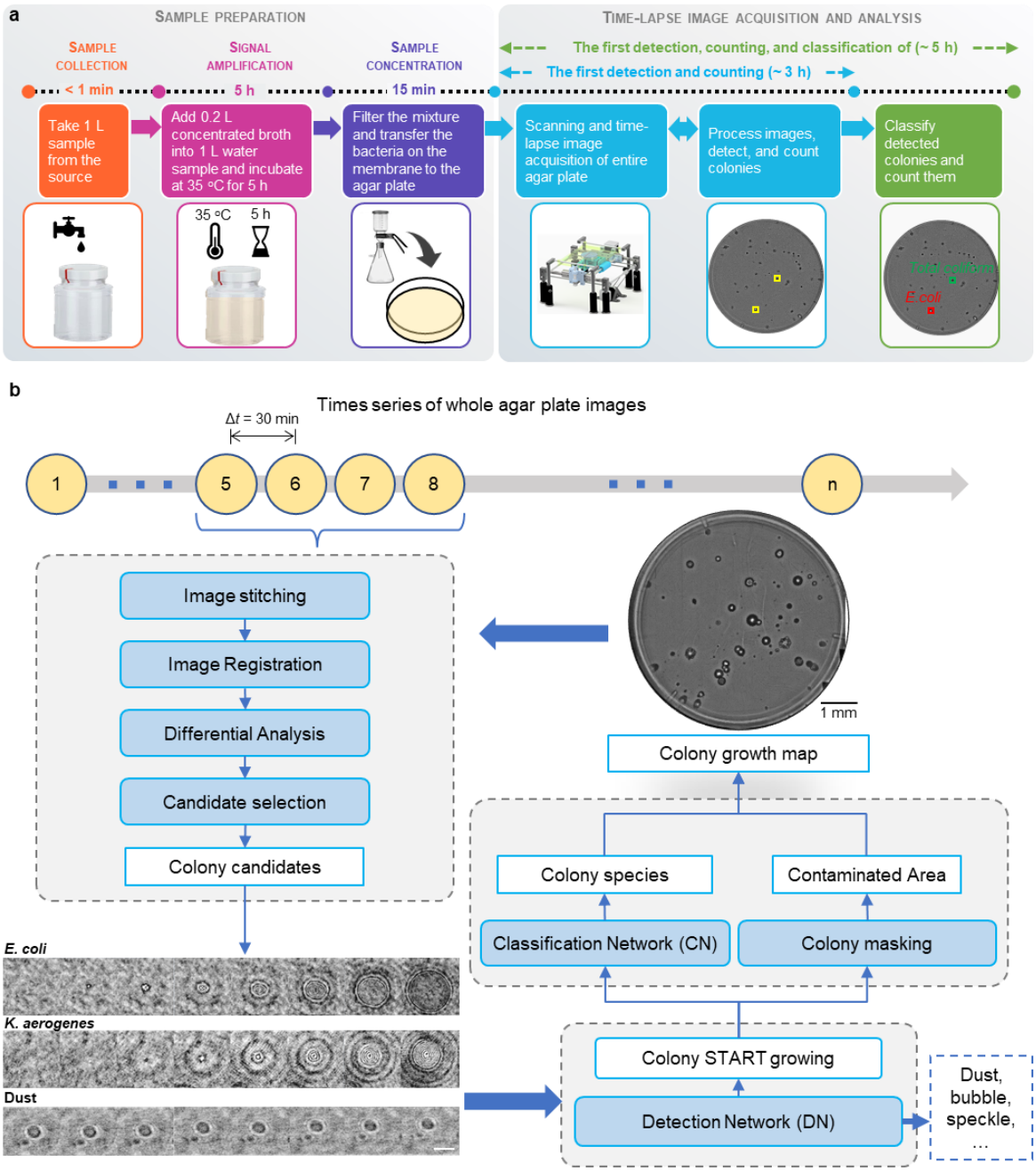

Figure 4.2 Schematics demonstrating the workflow of the microorganism monitoring system. (a) Bacterial sample preparation workflow. (b) Steps of the image and data processing algorithms for the automated detection of the growing colonies and classification of their species. The scale bars for the holographic images of the growing colonies (*E. coli* and *K. aerogenes*) and a static particle (dust) are 100 µm.

4.2 Design and training of neural networks for bacterial growth detection and classification

I designed a two-step framework for bacterial growth detection and classification. The first step selects colony candidates with differential image analysis and refines the results with a detection DNN. I designed a pseudo-3D (P3D) DenseNet [210] architecture to process the complex-valued (i.e., phase and amplitude) time-lapse image stacks (see the Methods section). In each time-lapse imaging experiment, I used 4 time-consecutive frames ($4 \times 0.5 = 2$ h) as a running window for the differential image analysis to extract individual regions of interest (ROIs) containing objects that changed their amplitude and/or phase signatures as a function of time. These initially detected objects that were extracted by the differential analysis algorithm were either growing colonies or surface impurities, e.g., from spreading the sample on the agar surface, evaporation of air bubbles in the agar plate, or coherent light speckles. I then used a DNN-based detection model to eliminate the non-bacterial objects and only kept the growing colonies (i.e., the true positives), as illustrated in **Figure 4.2**b. I used sensitivity (or true positive rate, TPR) and precision (or positive predictive value, PPV) measurements to quantify the results. Sensitivity is defined as:

$$\mathrm{TPR} = \mathrm{TP} / \mathrm{P},$$

where TP refers to the number of true positive predictions from this system, and P refers to the total number of colonies resulting from manual plate counting *after* 24 h (i.e., the ground truth). Precision is defined as:

$$\mathrm{PPV} = \mathrm{TP} / (\mathrm{TP} + \mathrm{FP}),$$

where FP refers to the number of false positive predictions from this system.

In total, 13,712 growing colonies (*E. coli*, *K. aerogenes*, and *K. pneumoniae*) and 30,000 non-colony objects captured from 66 separate agar plates were used in the training phase. Another 2,597 colonies and 13,078 non-colony objects from 5 independent plates were used as validation dataset to finalize the network models and achieved a TPR of ~95% and a PPV of ~95% once the network converged, which took ~ 4 h of training time.

The second step further classifies the species of the detected colonies with a classification DNN model following a similar network architecture. To accommodate the different growth rates of bacterial colonies, I used a longer time window in this classification neural network, containing 8 consecutive frames (8 × 0.5 = 4 h) for each sub-ROI. Since the bacterial growth detection network uses a shorter running time window of 2 h, there is a natural 2-h time delay between the successful detection of a growing colony and the classification of its species. The network was trained with 7,919 growing colonies, which contained 3,362 *E. coli*, 1,880 *K. aerogenes*, and 2,677 *K. pneumoniae* colonies, and it was validated with 340 *E. coli*, 205 *K. aerogenes*, and 988 *K. pneumoniae* colonies from 6 independent plates and reached a validation classification accuracy of ~89% for *E. coli*, ~95% for *K. aerogenes*, and ~98% for *K. pneumoniae* when the network model converged.

After these network models were finalized through the training and validation data, I tested their generalization capabilities with an additional set of experiments that were never seen by the networks before; the results of these blind tests are detailed next.

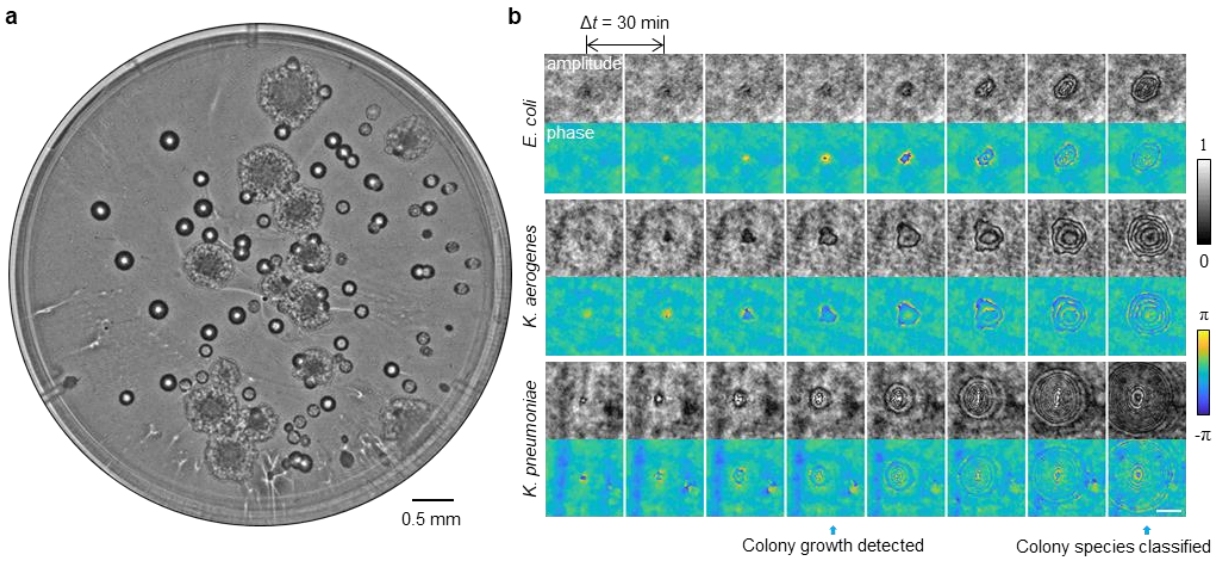

Figure 4.3 Images captured using the microorganism monitoring system. (a) Whole agar plate image of mixed *E. coli* and *K. aerogenes* colonies after 23.5 h of incubation. (b) Example images (i.e., amplitude and phase) of the individual growing colonies detected by a trained deep neural network. The time points of detection and classification of growing colonies are annotated with blue arrows. The scale bar is 100 μm.

4.3 Blind testing results for the early detection of bacterial growth

First, the performance of this system in the early detection of bacterial colonies was blindly tested with 965 colonies from 15 plates that were not presented during the network training or validation stages. I compared the predicted number of growing colonies on the sample within the first 14 h of incubation against a ground truth colony count obtained from plate counting *after* 24 h of incubation time. Each of the 3 sensitivity curves (**Figure 4.4**a-c) were averaged across repeated experiments for the same species, e.g., 4 experiments for *K. pneumoniae*, 7 experiments for *E. coli*, and 4 experiments for *K. aerogenes*, so that each data point was calculated from ~300 colonies. The results demonstrated that this system was able to detect 80% of the true positive colonies within ~6.0 h of incubation for *K. pneumoniae*, ~6.8 h of incubation for *E. coli*, and ~8.8 h of incubation for *K. aerogenes*. Additionally, this platform further detected 90% of the true

positives after ~1 additional hour of incubation and >95% of the true positive colonies of all 3 species within 12 h. The results also reveal that the early detection sensitivities in **Figure 4.4**a-c are dependent on the length of the lag phase of each tested bacteria species, which demonstrates inter-species variations. For example, *K. pneumoniae* started to grow earlier and faster than *E. coli* and *K. aerogenes*, whereas *K. aerogenes* did not reach a detectable growth size until 5 h of incubation. Furthermore, when the tails of the sensitivity curves were examined, some of the *E. coli* colonies showed late "wake-up" behaviour, as highlighted by the purple arrow in **Figure 4.4**b. Although most of the *E. coli* colonies were detected within ~10 h of incubation time, some of them did not emerge until ~11 h after the start of the incubation phase.

I also quantified the false positive rate of this platform with the PPV curve shown in **Figure 4.4**d, which was averaged across all the experiments covering all the species, i.e., 965 colonies from 15 agar plates. The precision can be low at the beginning of the experiments (the first 4 h of incubation) because the number of detected true positive colonies is very small, especially for *K. aerogenes*. This result means that even a single false positive-detected colony can dramatically affect the precision calculation. Nevertheless, the precision quickly rises up to ~100% within 6 h of incubation and is maintained at 99.2-100% for all the tested species after 7 h of incubation.

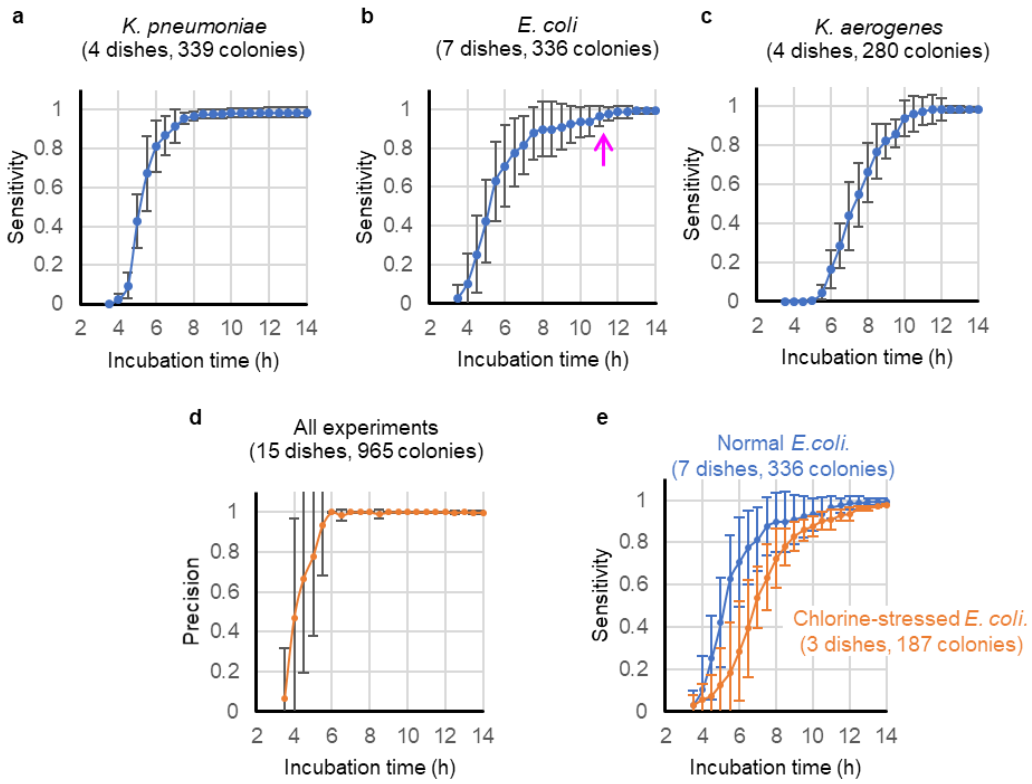

Figure 4.4 Sensitivity of growing colony detection using the trained neural network for (a) *K. pneumoniae*, (b) *E. coli*, and *K. aerogenes*. (d) Precision of growing colony detection using the trained neural network for all three species. The pink arrow indicates the time for late "wake-up" behavior for some of the *E. coli* colonies. (e) Characterizing the growth speed of chlorine-stressed *E. coli* using the system. There was an ~2 h delay in colony formation for chlorine-stressed *E. coli* (orange curve) compared to the unstressed *E. coli* strain (blue curve). The error bars show the standard deviation values across multiple plates.

I should emphasize here that the results presented in **Figure 4.4** represent the *lower limits* of the detection capabilities of this system since I calculated these sensitivities with regard to the number of true positive colonies *after* 24 h of incubation, whereas some of these colonies actually did *not* exist at the early stages due to delayed growth; stated differently in some cases, there were no colonies present at the early stages of the incubation period. I also note that the rising sensitivity curves in the results stand for the emergence of new bacterial colonies, in addition to the growth of colonies. Even though the sensitivity curves converge to flat lines after 12 h, the colonies

continue to grow exponentially until much later. Therefore, this system detects emerging colonies at an early stage, when they first appear, forming microscale features invisible to the naked eye.

These observations also indicate that this system can be very effective and used for high-throughput quantitative studies to better understand microorganism behaviour under different conditions, such as the evaluation of the differences in growth rates between stressed bacteria (e.g., under nutrient deprivation or chlorine treatment) and normal bacteria. [211–215] There are several reasons to detect and enumerate chlorine-stressed or injured coliform bacteria. First, the detection of injured *E. coli* or total coliform bacteria is directly related to the sensitivity of the detection platform. [215] For an effective and sensitive detection platform, false negative results should be avoided for public health safety. Another important reason is that the detection of injured *E. coli* or low numbers of *E. coli* in water samples is correlated with Salmonella outbreaks, a foodborne pathogen causing 1.2 million illnesses and ~500 deaths per year in the US [216], which forms an indirect indicator of contamination in irrigation water. [217] To evaluate the capabilities of this system to detect injured bacteria, 3 agar plates containing chlorine-stressed *E. coli* (see 4.6 Materials and methods) were prepared and imaged and characterized their growth using my detection workflow, as summarized in **Figure 4.4**e. The results indicate that this system can detect colony formation for chlorine-stressed *E. coli* on average with an ~2 h delay compared to the regular *E. coli* strain.

4.4 Blind testing results on the classification of growing bacteria

In addition to providing significant detection time savings while also achieving very good sensitivity and precision for the early detection of bacterial growth, the presented method also provides the automated classification of the corresponding species of the detected bacteria using a trained neural network. Therefore, an additional advantage of this system is its capability to further

classify the total coliform subspecies, which is not possible with traditional agar plate counting methods. For example, both *K. pneumoniae* and *K. aerogenes* colonies appear mauve in the agar plates. However, since the classification neural network not only relies on the byproducts of colorimetric reactions, it can successfully distinguish between different species based on their unique spatiotemporal growth signatures acquired by this platform at the microscale.

Figure 4.5 shows the blind testing results on species classification using the same experiments reported in the blinded early detection tests, containing 965 colonies of 3 different species from 15 agar plates. In these results, if a colony was not detected in the previous step (i.e., a false negative event compared to the 24 h reading), then it was naturally not sent to the classification neural network. The recovery rate was defined as the number of colonies correctly classified into their corresponding species using this system divided by the total number of colonies counted *after* 24 h. As the classification of each individual colony is an independent event, the recovery rate for each bacterial species (reported in **Figure 4.5**a-c) was calculated using all of the colonies detected in the previous step, i.e., 336, 280, and 339 colonies of *E. coli*, *K. aerogenes*, and *K. pneumoniae*, respectively. The shaded area in each curve represents the *highest* and *lowest* recovery rates found in all the corresponding experiments at each time point. The classification neural network correctly classified ~80% of all of the colonies within ~7.6 h, ~8 h, and ~12 h for *K. pneumoniae*, *E. coli*, and *K. aerogenes*, respectively. I once again emphasize that the results presented in **Figure 4.5**a-c represent the lower limits of the classification capabilities of this system since ground truth is acquired after 24 h of incubation. In reality, at various earlier time points within the incubation period, there was no growth for certain regions of the plates, which exhibited significantly delayed growth. To further demonstrate the classification performance of the trained neural network in a manner that is decoupled from the sensitivity of the previous detection network, the classification

confusion matrix is report in **Figure 4.5**d for all the colonies that were sent to the classification network for blind testing at 12 h after the start of the incubation. The trained network achieved classification accuracies of ~97.2%, ~84.0%, and ~98.5% for *E. coli*, *K. aerogenes*, and *K. pneumoniae*, respectively.

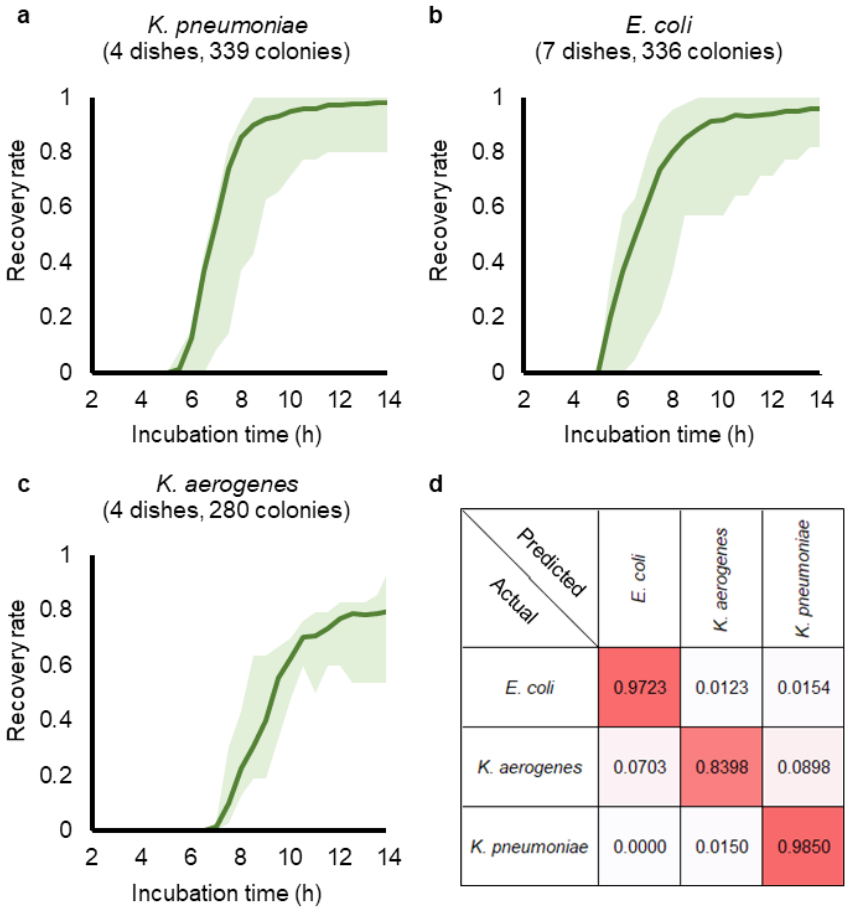

Figure 4.5 Classification performance of our trained neural network for (a) *K. pneumoniae*, (b) *E. coli*, and (c) *K. aerogenes* colonies. The green shaded area in each curve represents the highest and lowest recovery rates found in all the corresponding experiments at each time point. (d) The blind testing confusion matrix of classifying all the colonies that were sent to our trained neural network after 12 h of incubation. A diagonal entry of 1.0 means a 100% classification accuracy for that species. The number of colonies that were tested by the classification network in (d): 325 (*E. coli*), 334 (*K. pneumoniae*), and 256 (*K. aerogenes*).

4.5 Limit of detection as a function of the total test time

The detection limit of this system was further quantified and compared its performance against both Colilert® 18, which is an EPA-approved method, and traditional plate counting (**Table 4.1** and **Error! Reference source not found.**). To compensate for the CFU loss during the sample transfer from the water suspension to the filter membrane, a signal amplification step was introduced by preincubating the water sample under test, mixing it with a growth medium for 5 h at 35 °C before the filtration step (see the Methods section for details). For each measurement, 2 agar plates were prepared and monitored at the same time for comparison, one of which was for the sample amplified with a 5-h preincubation step before filtering, while the other was for the sample directly filtered and transferred to the agar plate (**Error! Reference source not found.**). B oth plates were incubated for the same amount of time at each imaging time point to provide a fair comparison between the two. The measurements were repeated using different concentrations of *E. coli* suspensions; these concentrations were compared to the average of three replicates of the same samples prepared using the Colilert®-18 method (**Error! Reference source not found.**). As shown in **Figure 4.7**a, this system is able to surpass the sensitivity of Colilert®-18 within ~8 h in total (including the time for signal amplification, sample concentration, and time-lapse imaging, altogether) and reach > 2 times the sensitivity of Colilert®-18 in ~9 h. I also quantified the LOD of this system by preparing and imaging 3 agar plates without bacteria, which show on average < 1 CFU count from the setup throughout the test period from 5 h to 14.5 h (**Figure 4.7**c), revealing a detection limit of $\mu + 3\sigma$ = ~2 CFU per test, where μ and σ refer to the mean and standard deviation of the detected CFU count, respectively. Due to the effective signal amplification enabled by the preincubation step, even with the lowest bacterial concentration of ~1 CFU/L, this system was able to detect 2 CFU at 8.5 h and 12 CFU at 9 h; in comparison, for the same

contaminated water sample, Colilert® 18 achieved 1.4±1.6 CFU/L after 18 h of incubation. Furthermore, for all the concentrations I have experimented with (~1-160 CFU/L), the system successfully detected more than 2 CFU per test in ≤ 9 h of test time, including all the necessary steps, i.e., the time for signal amplification, sample concentration, and time-lapse imaging; these results reveal that this system with a preincubation step achieves a detection limit of ~1 CFU/L within ≤ 9 h of total test time.

I also observe in **Figure 4.7**b that without the signal amplification enabled by preincubation, the detection performance is negatively affected due to the low transfer rate of bacteria from the container to the agar plate. In general, the sensitivity and LOD of this method might be further improved by increasing the preincubation time of the water-broth mixture at the cost of an increase in the total time to achieve automated detection and classification.

Table 4.1 Colony counts of some *E. coli* spiked samples in comparison to Colilert®-18 and plate counting.

Colilert®-18					Plate counting (TSA plates)					Plate counting (ECC ChromoSelect Selective Agar plates)				
R1*	R2*	R3*	Average	Std. deviation	R1†	R2†	R3†	Average	Std. deviation	R1†	R2†	R3†	Average	Std. deviation
172.3	172.3	135.4	160.00	21.30	169	162	198	176.33	19.09	164	137	140	147.00	14.80
11	17.3	20.1	16.13	4.66	15	18	14	15.67	2.08	17	13	17	15.67	2.31
225.4	166.4	228.2	206.67	34.90	228	260	246	244.67	16.04	245	241	221	235.67	12.86
8.6	8.5	12.1	9.73	2.05	4	4	5	4.33	0.58	2	5	11	6.00	4.58
37.9	43.5	32.3	37.9	5.6	52	37	30	39.67	11.24	35	28	36	33.00	4.36
3.1	1	<1	2.05	1.48	3	1	0	1.33	1.53	3	3	2	2.67	0.58
107.6	113.7	101.7	107.67	6.00	76	116	99	97.00	20.07	150	134	123	135.67	13.58
172.3	210.5	121.1	167.97	44.86	165	165	141	157.00	13.86	169	171	164	168.00	3.61

R is for replicate sample

* CFU per 100 mL

† CFU per 0.1 mL

Table 4.2 Comparison of our device against a scanning bright-field microscope for imaging of an agar plate (60 mm diameter).

Configuration	This work	Bright-field microscope (4 ×/0.1 NA objective lens)
Field of view (FOV) per image (mm^2)	29.4	14.4
Total FOV scanned (mm^2)	~3491	~2977
Total imaging time per agar plate (min)	~1.5	128
Effective pixel count (million)	570	435
Observation depth (μm)	> 20,000	3,000 (with 20 μm accuracy) *

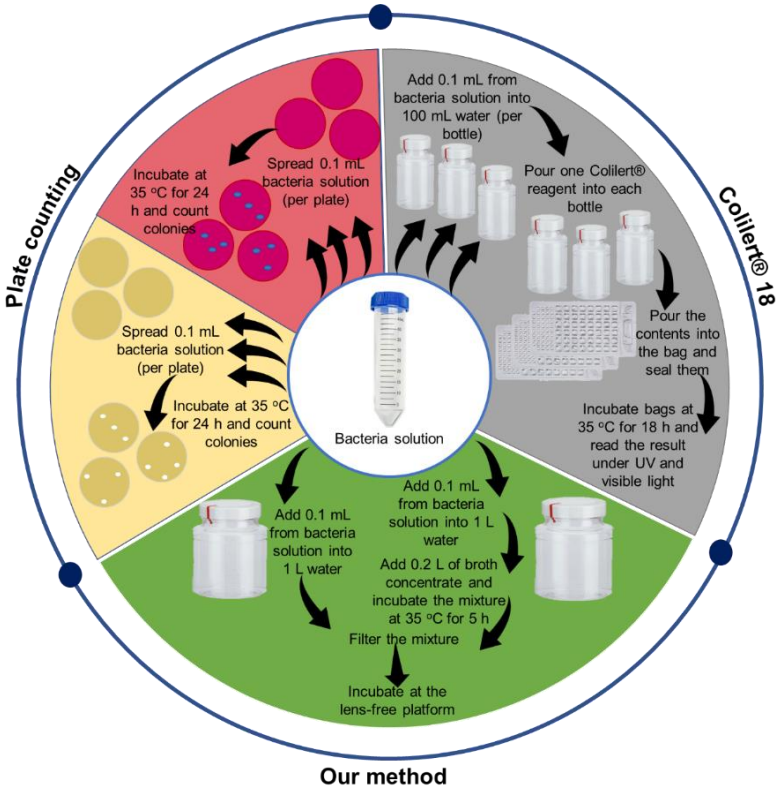

Figure 4.6 Schematics comparing the major steps involved in each one of the three different methods analyzed in this work.

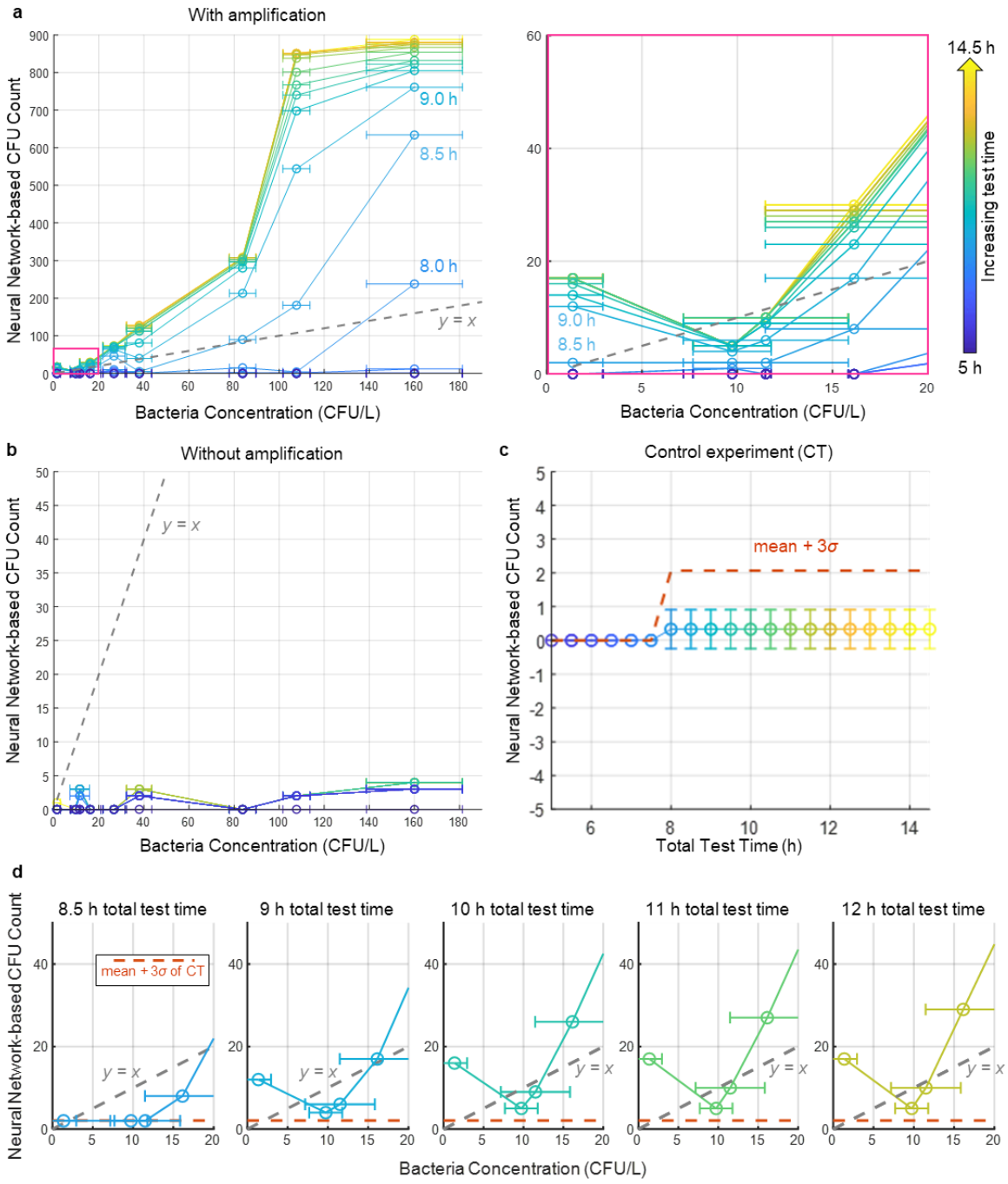

Figure 4.7 Quantification of the LOD of the presented system. (a) The CFU count from the system is plotted against the CFU/L counts of the spiked samples, calculated independently using the Colilert®18 method after 18 h of incubation. CFU counts acquired with this platform at different time points are coloured from blue to yellow, which corresponds to 5 to 14.5 h of total test time, including the signal amplification step that involves liquid culture media (5 h). (b) Without signal amplification, the LOD is decreased due to the low transfer rate from the filter membrane to

the agar surface (see **Figure 4.6** and **Figure 4.8**). (c) As a control experiment, 3 agar plates were prepared and imaged that showed < 1 CFU count from this setup throughout the test period from 5 h to 14.5 h. (d) The LOD of the system is ~11 CFU/L at 8.5 h and ~1 CFU/L at ≤9 h.

4.6 Materials and methods

Sample preparation

Safety practices

All the bacterial cultures and experiments were handled and performed at a Biosafety Level 2 laboratory in accordance with the environmental, health, and safety rules of the University of California, Los Angeles.

Studied organisms

E. coli (Migula) Castellani and Chalmers (ATCC® 25922™) (risk level 1), *K. aerogenes* Tindall *et al.* (ATCC® 49701™) (risk level 1), and *K. pneumoniae* subsp. *pneumoniae* (Schroeter) Trevisan (ATCC®13883™) (risk level 2) were used as the culture organisms.

Preparation of the poured agar plates

CHROMagar™ ECC (product no. EF322, DRG International, Inc., Springfield, NJ, USA) chromogenic substrate mixture was used as the solid growth medium for the detection of *E. coli* and total coliform colonies. CHROMagar™ ECC (8.2 g) was mixed with 250 mL of reagent grade water (product no. 23-249-581, Fisher Scientific, Hampton, NH, USA) using a magnetic stirrer bar. The mixture was then heated to 100 °C on a hot plate while being stirred regularly. After cooling the mixture to ~50 °C, 10 mL of the mixture was dispensed into Petri dishes (60 mm × 15 mm) (product no. FB0875713A, Fisher Scientific, Hampton, NH, USA). The agar plates were allowed to solidify, were sealed using parafilm (product no. 13-374-16, Fisher Scientific, Hampton,

NH, USA), and were covered with aluminum foil to keep them in the dark before use. The plates were stored at 4 °C and were used within two weeks of preparation.

Preparation of the melted agar plates

CHROMagar™ ECC (3.28 g) was mixed with 100 mL of reagent grade water using a magnetic stirrer bar, and the mixture was heated to 100 °C. After the mixture cooled to ~40 °C, 1 mL of the bacterial suspension was mixed with the agar and dispensed into Petri dishes. The plates were either incubated in a benchtop incubator (product no. 51030400, ThermoFisher Scientific, Waltham, MA, USA) or in the presented imaging platform (for monitoring the bacterial growth digitally). Tryptic soy agar was used to culture *E. coli* at 37 °C and *K. aerogenes* at 35 °C and nutrient agar to culture *K. pneumoniae* at 37 °C. Twenty grams of tryptic soy agar (product no. DF0369-17-6, Fisher Scientific, Hampton, NH, USA) or 11.5 g of nutrient agar (product no. DF0001-17-0, Fisher Scientific, Hampton, NH, USA) were suspended in 500 mL of reagent grade water using a magnetic stirrer bar. The mixture was boiled on a hot plate and then autoclaved at 121 °C for 15 min. After the mixture cooled to ~50 °C, 15 mL of the mixture was dispensed into Petri dishes (100 mm × 15 mm) (product no. FB0875713, Fisher Scientific, Hampton, NH, USA), which were then sealed with parafilm and covered with aluminum foil to keep them in the dark before use. The Petri dishes were stored at 4 °C until use.

Preparation of the chlorine-stressed E. coli samples

E. coli grown on tryptic soy agar plates and incubated for 48 h at 37 °C was used. Disposable centrifuge tubes (50 mL) were used as a sample container, and the sample size was 50 mL. Five hundred milliliters of reagent grade water was filtered for sterilization using a disposable vacuum filtration unit (product no. FB12566504, Fisher Scientific, Hampton, NH, USA). A fresh chlorine

suspension was prepared in a 50 mL disposable centrifuge tube to a final concentration of 0.2 mg/mL using sodium hypochlorite (product no. 425044, Sigma Aldrich, St. Louis, MO, USA), mixed vigorously, and covered with aluminum foil. [218] Sodium thiosulfate (10% [w/v]) (product no. 217263, Sigma Aldrich, St. Louis, MO, USA) in reagent grade water was prepared, and 1 mL of the solution was filtered using a sterile disposable syringe and a syringe filter membrane (product no. SLGV004SL, Fisher Scientific, Hampton, NH, USA) for sterilization. Water suspensions were prepared by spiking *E. coli* into filtered water samples. Fifty microliters of the chlorine suspension (i.e., 0.2 ppm) was added to the test water sample, and a timer counted the chlorine exposure time. The reaction was stopped at 10 min of chlorine exposure by adding 50 µL sodium thiosulfate into the test water sample and vigorously mixing the solution to immediately stop the chlorination reaction. CHROMagar™ ECC plates were inoculated with 200 µL of the chlorine-stressed suspension, were dried in the biosafety cabinet for at most 30 min and then were placed on the setup for lens-free imaging. In addition, three TSA plates and one ECC ChromoSelect Selective Agar plate (product no. 85927, Sigma Aldrich, St. Louis, MO, USA) were inoculated with 1 mL of the control sample (not exposed to chlorine) and 0.2 ppm of the chlorine-stressed *E. coli* water sample and dried under a biosafety cabinet for approximately 1-2 h with the gentle mixing of Petri dishes at some time intervals. After drying, the plates were sealed with parafilm and incubated at 37 °C for 24 h. After incubation, the bacterial colonies grown on the agar plates were counted, and the *E. coli* concentrations of the control samples and chlorine-stressed *E. coli* samples were compared. If the achieved reduction in colony count was between 2.0-4.0 log, then the images of CHROMagar™ ECC plates captured using the lens-free imaging platform were used for further analysis.

Preparation of the culture plates for lens-free imaging

A bacterial suspension in a phosphate buffer solution (PBS) (product no. 20-012-027, Fisher Scientific, Hampton, NH, USA) was prepared every day from a solid agar plate incubated for 24 h. The concentration of the suspension was measured using a spectrophotometer (model no. ND-ONE-W, Thermo Fisher), and the suspension was then diluted in PBS to a final concentration of 1–200 CFU per 0.1 mL. One hundred microliters of the diluted suspension was spread on a CHROMagar™ ECC plate using an L-shaped spreader (product no. 14-665-230, Fisher Scientific, Hampton, NH, USA). The plate was covered with its lid, inverted, and incubated at 37 °C in the presented optical platform (**Figure 4.2**).

Preparation of a concentrated broth

A total of 180 g of tryptic soy broth (product no. R455054, Fisher Scientific, Hampton, NH, USA) was added to 1 L reagent grade water and heated to 100 °C by continuously mixing using a stirrer bar. The suspension was then cooled to 50 °C and filter sterilized using a disposable filtration unit. The broth concentrate was stored at 4 °C and used within one week after preparation.

Preparation of samples for comparison measurements

The performance of the presented method was evaluated in comparison to Colilert® 18, which is an EPA-approved enzyme-based analytical method for several types of regulated water samples (e.g., drinking water, surface water, ground water) to detect *E. coli* [219] and for plate counting using TSA plates and ECC ChromoSelect Selective Agar plates (**Error! Reference source not found.**). Two bottles of 1 L reagent grade water were filtered using disposable vacuum filtration units and 0.2 L of the concentrated broth was added into one of the 1 L sample bottles. The bottles were covered with aluminum foil and stored in a biosafety cabinet overnight. A glass vacuum filtration unit was used for the filtration of the 1 L water samples. The components of the unit were

covered with aluminum foil and sterilized using an autoclave. The disposable nitrocellulose filter membranes (product no. HAWG04705, EMD Millipore, Danvers, MA, USA) used in the glass filtration unit were also sterilized using the autoclave. A bacterial suspension was prepared by spiking bacteria into 50 mL reagent grade water using a disposable inoculation loop from a TSA plate containing *E. coli* colonies. The suspension was mixed gently to obtain a uniform distribution of bacteria. Three TSA plates, 3 ECC ChromoSelect Selective Agar plates, and 4 CHROMagar™ ECC plates were removed from the refrigerator and were kept at room temperature for 30 min.

Three bottles of 120 mL disposable vessels with sodium thiosulfate (product no. WV120SBST-200, IDEXX Laboratories Inc., Westbrook, ME, USA) were filled with 100 mL filter sterilized reagent grade water. First, 0.1 mL of bacterial suspension was spiked into a 1 L water sample, a 1.2 L water sample (1 L water + 0.2 L concentrated broth), 3 bottles of 100 mL water samples, 3 TSA plates and 3 ECC ChromoSelect Selective Agar plates, sequentially. The timer was started immediately after adding the spike into the suspensions.

First, the suspensions on TSA plates and ECC ChromoSelect Selective Agar were spread using L-shaped disposable spreaders. Then, the water sample with broth was mixed for approximately one minute and then stored at 35 °C for 5 h. One Colilert® 18 reagent (product no. 98-27164-00, IDEXX Laboratories Inc., Westbrook, ME, USA) was added into each 100 mL bacterial suspension, and the mixture was shaken. The content of the bottle was poured into a Quanti-Tray 2000 bag (product no. 98-21675-00, IDEXX Laboratories Inc., Westbrook, ME, USA), and after removing bubbles in each well, the bag was sealed using Quanti-Tray Sealer (product no. 98-09462-01, IDEXX Laboratories Inc., Westbrook, ME, USA). Three bags sealed and labelled with the experimental details were incubated at 35 °C for 18 h. Next, 30 mL filtered reagent grade water was used to moisturize the membrane in the glass filtration unit, and then an

E. coli-contaminated 1 L water sample was filtered at a pressure of 50 kPa. The bottle was rinsed using 150 mL of sterilized reagent grade water, and the solution was filtered on the unit (**Figure 4.8**). The funnel was rinsed twice using 50 mL of sterilized reagent grade water. After the filtration was complete, the membrane was removed and placed onto a CHROMagar™ ECC plate face down. Gentle pressure was applied on the membrane using a tweezer to remove any air bubbles between the agar and the membrane. Then, 30 g of weight was placed on the membrane to provide continuous pressure during the transfer of bacteria from the membrane to the agar plate. After 5 min of incubation, the membrane was gently peeled off from the agar surface and placed into another agar facing up. The agar containing the membrane was incubated at the benchtop incubator at 35 °C, and the agar containing the transferred bacteria was incubated at the lens-free imaging platform for time-lapse imaging. After 5 h of incubation, the bottle containing 1.2 L suspension was filtered using the same procedure as described before for filtration of a 1 L sample. The agar plate containing the transferred bacteria was incubated at the second sample tray of the lens-free imaging setup for time-lapse imaging, while the agar containing the membrane was incubated at the benchtop incubator.

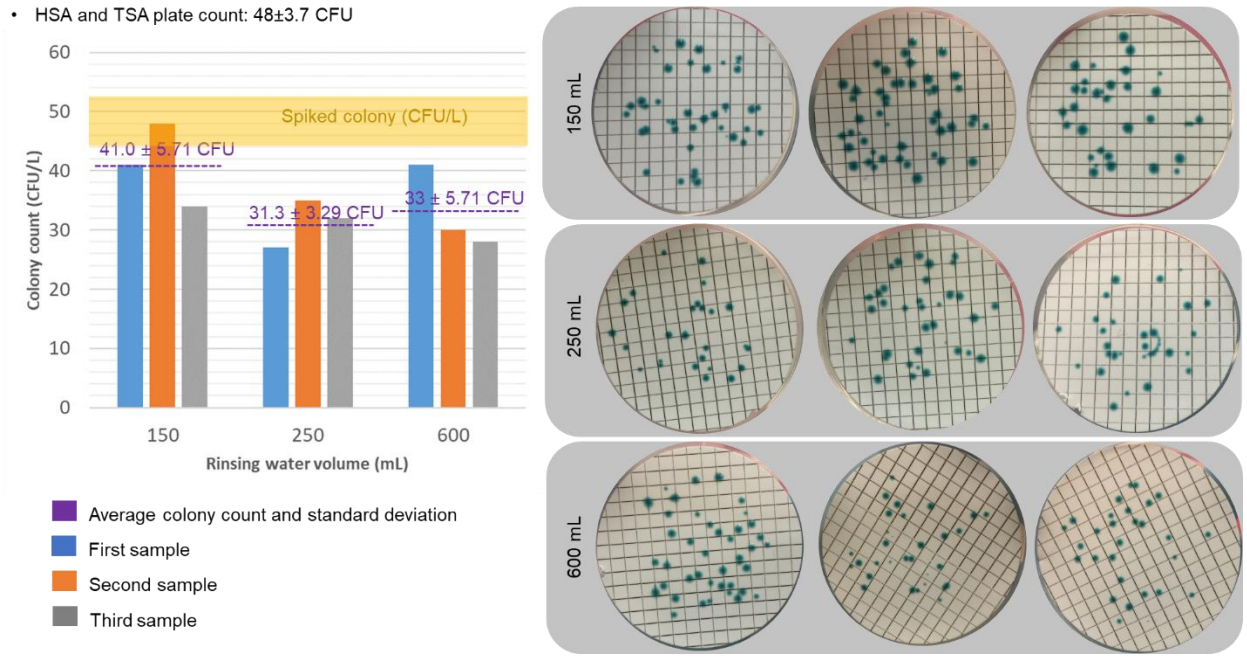

Figure 4.8 Colony counts obtained for optimization of the amount of water used for washing the sample container.

Design of the high-throughput time-resolved microorganism monitoring platform

The presented platform consists of five modules: (1) a holographic imaging system, (2) a mechanical translational system, (3) an incubation unit, (4) a control circuit, and (5) a controlling program. Each module is explained in detail below.

- A fiber-coupled partially coherent laser illumination (SC400-4, Fianium Ltd, Southampton, UK) was used, with the wavelength and intensity controlled through an acousto-optic tunable filter (AOTF) device (Fianium Ltd, Southampton, UK). The device was remotely controlled with a customized program written in the C++ programming language and ran on a controlling laptop computer (product no. EON17-SLX, Origin PC). The laser light was transmitted through the sample, i.e., the agar plate that contains the bacterial colonies, and forms an inline hologram on a CMOS image sensor (product no. acA3800-14 μm, Basler AG, Ahrensburg, Germany) with a pixel size of 1.67 μm and an active area of 6.4 mm × 4.6

mm. The CMOS image sensor was connected to the same controlling laptop computer through a universal serial bus (USB) 3.0 interface and was software-triggered within the same C++ program. The exposure time at each scanning position was precalibrated according to the intensity distribution of the illumination light and ranged from 4 ms to 167 ms. The images were saved as 8-bit bitmap files for further processing.

- The mechanical stage was customized with a pair of linear translation rails (Accumini 2AD10AAAHL, Thomson, Radford, VA, USA), a pair of linear bearing rods (8 mm-diameter, generic), and linear bearings (LM8UU, generic), and it was aided by parts printed by a 3D printer for the joints and housing (Objet30 Pro, Stratasys, Minnesota, USA). The 2D horizontal movement was powered by two stepper motors (product no. 1124090, Kysan Electronics, San Jose, CA, USA)—one for each direction, and these motors were individually controlled using stepper motor controller chips (DRV8834, Pololu Las Vegas, NV, US). To minimize the backslash effect, the whole Petri dish was scanned following a raster scan pattern.

- The incubation unit was built with the top heating plate of a microscope incubator (INUBTFP-WSKM-F1, Tokai Hit, Shizuoka, Japan), and it was housed by a 3D frame printed by a 3D printer. The Petri dish containing the sample was placed on the heating plate with the surface having bacteria facing downwards. The temperature was controlled by a paired controller that maintained a temperature of 47 °C on the heating plate, resulting in a temperature of 38 °C inside the Petri dish.

- The control circuit consisted of three components: a microcontroller (Arduino Micro, Arduino LLC) communicating with the computer through a USB 2.0 interface, two stepper motor driver chips (DRV8834, Pololu Las Vegas, NV, US) externally powered by a 4.2 V

constant voltage power supply (GPS-3303, GW Instek, Montclair, CA, US), and a metal–oxide–semiconductor field-effect transistor (MOSFET)-based digital switch (SUP75P03-07, Vishay Siliconix, Shelton, CT, United States) for controlling the CMOS sensor connection.

- The controlling program included a graphical user interface (GUI) and was developed using the C++ programming language. External libraries including Qt (v5.9.3), AOTF (Gooch & Housego), and Pylon (v5.0.11) were integrated.

Data acquisition

Inoculated agar plates of pure bacterial colonies were prepared and captured images of an entire agar plate at 30-min intervals. The illumination light was set to a wavelength of 532 nm and an intensity of ~400 μW. To maximize the image acquisition speed, the captured images were first saved into a computer memory buffer and then were written to a hard disk by another independent thread. At the end of each experiment (i.e., after 24 h of incubation), the sample plate was imaged using a benchtop scanning microscope (Olympus IX83) in reflection mode, and the resulting images were automatically stitched to a full-FOV image, used for comparison. Subsequently, the plate was disposed of as solid biohazardous waste. The data (i.e., time-lapse lens-free images) was populated corresponding to ~6,969 *E. coli*, ~2,613 *K. aerogenes*, and ~6,727 *K. pneumoniae* individual bacterial colonies to train and validate the models. Another 965 colonies of 3 different species from 15 independent agar plates were used to blindly test the machine learning models.

Image processing and analysis

The acquired lens-free images were processed using custom-developed image processing and deep learning algorithms. Five major image processing steps were used for the early detection and automated classification and counting of colonies. These steps are described in detail below.

Image stitching to obtain the image of the entire plate area:

Following the acquisition of holographic images using the multi-threading approach, all the images within a tile-scan of the whole Petri dish per wavelength were merged into a single full-FOV image. During a tile scan, the images were acquired with ~30% overlap on each side of the image to calculate the relative image shifts against each other. For each image, the relative shifts against all four of the neighbouring images were calculated using a phase correlation [132] method, followed by an optimization step that minimized an object function, as defined by:

$$\arg\min_{T_{VF}} \sum_{A \in V \setminus \{F\}} \left(\sum_{B \in V \setminus \{F\}} \left\| \vec{t}_{AF} - \vec{t}_{BF} - \vec{p}_{AB} \right\|^2 \right), \tag{4.1}$$

where V is the set of all tile images, $F \in V$ is a fixed image, e.g., the image captured at the centre of the sample Petri dish, $\vec{t}_{AB}$ stands for the relative position of image A with respect to image B, and $\vec{p}_{AB}$ is the local shift between images A and B, calculated by the phase correlation method using the overlapping regions of the two neighbouring images, which can be formulated as:

$$\vec{p}_{AB} = (\Delta x, \Delta y) = \arg\max_{(x,y)} \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{A\} \cdot \mathcal{F}\{B\}^*}{\left| \mathcal{F}\{A\} \cdot \mathcal{F}\{B\}^* \right|} \right\} \tag{4.2}$$

where $\mathcal{F}$ is the Fourier transform operator and $\mathcal{F}^{-1}$ is the inverse Fourier transform operator. The optimal configuration $T_{VF} = \{\vec{t}_{AF} : A, F \in V\}$ represents the relative positions of all the images with respect to the fixed image F, and it was used as the global position of each tile image for full-FOV image stitching. To eliminate tiles with a low signal-to-noise ratio (SNR) that lead to incorrect local shift estimation values, a correlation threshold of 0.3 was applied during the optimization, meaning that if the cross-correlation coefficient of the overlapped parts of two images was below 0.3, the shift calculation was discarded. Once the positions of all of the tiles were obtained, they

were merged into a full-FOV image of the whole Petri dish using linear blending. A full-FOV image of the whole Petri dish was defined as a 'frame'. All the frames were normalized so that the mean value was 50, and they were saved as unsigned 8-bit integer (0-255) arrays.

Colony candidate selection by differential analysis:

When a new frame was acquired at time *t*, it was cross-registered to the previous frame at time $t-1$ and then digitally back-propagated to the sample plane [15,220] to obtain the complex light field

$$\tilde{B}_t = \mathrm{P}(F_t, \mathbf{z}), \tag{4.3}$$

where F_t is the frame at time t, $\boldsymbol{z}$ is a surface normal vector of the sample plane obtained by digital autofocusing [221] at 50 randomly spaced positions, and P denotes the angular spectrum-based back-propagation operation, [15,220] which can be calculated by multiplying the spatial Fourier transform of the input signal and the following transfer function:

$$H_k(\nu_x, \nu_y) = \begin{cases} \exp\left[-j \cdot 2\pi \dfrac{n \cdot z}{\lambda} \sqrt{1-\left(\dfrac{\lambda}{n}\nu_x\right)^2 - \left(\dfrac{\lambda}{n}\nu_y\right)^2}\right] & \left(\nu_x^2 + \nu_y^2 \le \left(\dfrac{n}{\lambda}\right)^2\right) \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

where *n* is the refractive index of the medium, λ is the illumination wavelength, and ν_x and ν_y are the spatial frequencies. This operation was followed by an inverse 2D Fourier transform. The resulting complex-valued reconstruction provides both the amplitude and phase images of the illuminated objects. To accommodate the large FOV of a stitched frame (36000 × 36000 pixels), digital back-propagation was performed with 2048 × 2048-pixel blocks, which were then merged together.

Four consecutive frames were taken, i.e., from $t-3$ to t, and a differential image was calculated defined by:

$$D_t = \mathrm{HP}\left[\mathrm{LP}\left(\frac{1}{3}\sum_{\tau=t-2}^{t}\left|\tilde{B}_\tau - \tilde{B}_{\tau-1}\right|\right)\right], \tag{4.5}$$

where D_t is the differential image at time *t*, $\tilde{B}_t$ represents the complex light field obtained by back-propagating frame *t*, and LP and HP represent low-pass and high-pass image filtering, respectively. The HP filter removes the differential signal from a slowly varying background (unwanted term), and the LP filter removes the high-frequency noise-introduced spatial patterns. The LP and HP filter kernels were empirically set to 5 and 100, respectively.

Following the differential image calculation, regions in the differential image with > 50 connective pixels that are above an intensity threshold were selected, which was empirically set to 12. These regions are marked as colony candidates, as they give a differential signal over a period of time (covering four consecutive frames). However, some of the differential signals come from non-bacterial objects, such as a water bubble or surface movement of the agar itself. Therefore, two DNNs to were also used select the true candidates and classify their species.

DNN-enabled detection of growing bacterial colonies

Following the colony candidate selection process outlined earlier, I cropped out candidate regions of 160 × 160 pixels (~267 μm × 267 μm) across the four back-propagated consecutive frames and separated the complex field into amplitude and phase channels. Therefore, each candidate region is represented by a 2 × 4 × 160 × 160 array. This four-dimensional (phase/amplitude-time-x-y) data format differs from the traditional three-dimensional data used in image classification tasks and requires a custom-designed DNN architecture that accounts for the

additional dimension of time. I designed the DNN by following the block diagram of DenseNet [210] and replaced the 2D convolutional layers with P3D convolutional layers [222], as shown in **Figure 4.9**. The network was implemented in Python (v3.7.2) with the PyTorch Library (v1.0.1). The network was randomly initialized and optimized using an adaptive moment estimation (Adam) optimizer [223] with a starting learning rate of 1×10^{-4} and a batch size of 64. To stabilize the accuracy of the network model, I also set a learning rate scheduler that decayed the learning rate by half every 20 epochs. Approximately 16,000 growing colonies and 43,000 non-colony objects captured from 71 agar plates were used in the training and validation phases. The best network model was selected based on the best validation accuracy. Data augmentation was also applied by random 90°-rotations and flipping operations in the spatial dimensions. The whole training process took ~5 h using a desktop computer with dual GPUs (GTX1080Ti, Nvidia). The decision threshold value after the softmax layer was set to 0.5 during training, i.e., positive for softmax value >0.5 and negative for softmax value <0.5, which implies equal penalty to false positive and false negative events. The threshold value was adjusted to 0.99, empirically based on the training dataset before blind testing, to favor fewer false positive events.

DNN-enabled classification of the bacterial colony species

Once the true bacterial colonies are selected, they grow for another 2 h to collect 8 consecutive frames, i.e., 4 h, and then are sent to the second DNN as a 2 × 8 × 288 × 288 array for the classification of colony species. To perform the classification task, this time, the training data only contain the true colonies and their corresponding species (ground truth). The network follows a similar structure and training process as the detection model, as illustrated in **Figure 4.9**. The network was randomly initialized and optimized using the Adam optimizer [223], with a starting learning rate of 1×10^{-4} and a batch size of 64. The learning rate decayed by 0.9 times every 10

epochs. To avoid overfitting to a specific plate, colony images extracted from extremely dense samples (>1000 CFU per plate) were discarded. As a result, approximately 9,400 growing colonies were used in the training and validation of the classification model. The whole training process took ~15 h using a desktop computer with dual GPUs (GTX1080Ti, Nvidia).

Colony counting:

The respective ground truth information on the growing colonies in each experiment was created after the sample was incubated for >24 h. At the boundary of the plate, the agar always forms a curved surface owing to surface tension, thereby distorting the images of the colonies. Therefore, the effective imaging area was limited to a 50 mm-diameter circle in the centre of the agar plate. In cases where multiple colonies are closely spaced and eventually merge into one large colony (e.g., towards the end of the 24 h incubation period), lens-free time-lapsed images were then used to verify the true colony number when detected by the presented method to avoid overcounting.

Calculation of the imaging throughput

In **Table 4.2**, the imaging throughput of the presented system was compared with a conventional lens-based scanning microscope in terms of the space-bandwidth product (SBP) [2] using the following formula:

$$N_{\mathrm{I}} = \alpha \cdot \mathrm{FOV} \cdot r^2 / \delta^2 \tag{4.6}$$

where N_{I} is the effective pixel count of a frame, δ is the half-pitch resolution, r is the digital sampling factor along the x and y directions, $\alpha = 2$ represents the independent spatial information contained in the phase and amplitude images of the holographic reconstruction, and $\alpha = 1$

represents the amplitude-only information contained in an image captured using the standard lens-based bright-field scanning microscope. In the lens-based microscope, I used a colour camera with a pixel size of 7.4 µm. Therefore, for a 4× objective lens, the image resolution is limited to ~3.7 µm, owing to the Nyquist sampling limit. Without loss of generality, I set $r = 2$. [51]

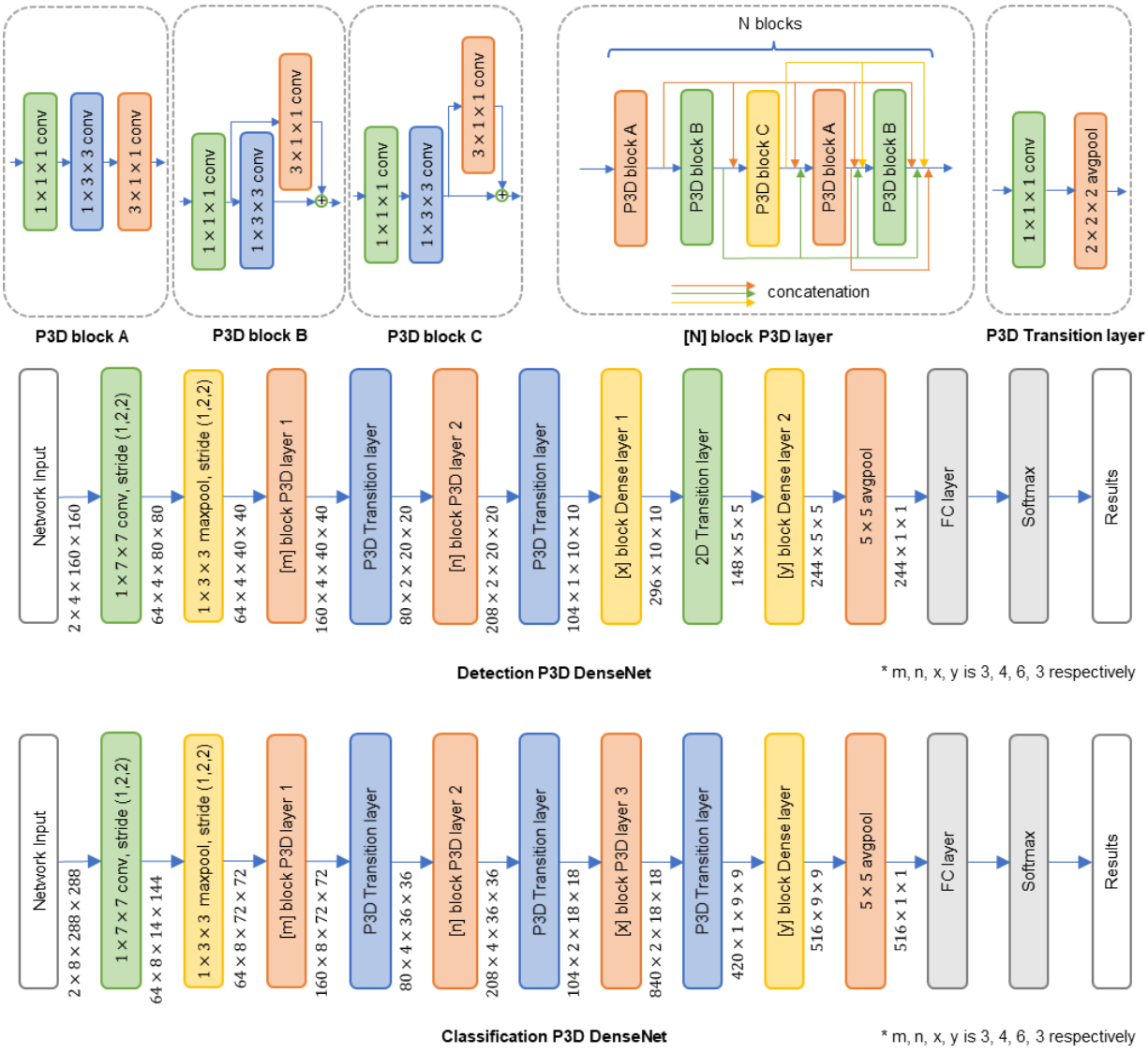

Figure 4.9 Schematic of pseudo-3D (P3D) DenseNet models for the detection and classification of growing colonies using the lens-free imaging system. The detection neural network model has 1.21×10^6 trainable parameters. The classification neural network model has 1.36×10^6 trainable parameters.

4.7 Discussion

A new platform for the early detection and classification of bacterial colonies was demonstrated, which is fully compatible with the existing EPA-approved methods and can be integrated with them to considerably improve the analysis of agar plates [224]. The presented approach can automatically detect bacterial growth as early as 3 h and can detect 90% of bacterial colonies within 7–10 h (and >95% within 12 h), with a precision of 99.2-100%. The system also correctly classifies ~80% of all of the tested bacterial colonies within 7.6, 8.8, and 12 h for *K. pneumoniae*, *E. coli*, and *K. aerogenes*, respectively. These results present a total time savings of more than 12 h compared to the gold-standard methods (e.g., Colilert test and Standard Method 9222 B), which require 18-24 h. The presented learning-based bacteria detection and classification framework can potentially be further advanced by training it with a larger number of sample types [202], and it can also be applied to other bacteria sensing applications beyond water quality monitoring. In addition to the automated detection of live bacteria and species classification, the rich spatiotemporal information embedded in the holographic images can be used for more advanced analysis of water samples and microbiology research in general.

Another advantage of this system is its high-throughput imaging capability of agar plates. The prototype performs a 242-tile scan within 87 s per agar plate, corresponding to a raw image scanning throughput of ~49 cm^2/min. To leave sufficient data redundancy for image postprocessing, I set a relatively large overlap of 30% on each side of the acquired holographic image, which reduces the effective imaging throughput of the platform to ~24 cm^2/min. As this system is based on lens-free holographic microscopy, it does not require mechanical axial focusing at each position and instead autofocuses onto the object plane computationally. The spatial resolution of this system was characterized by imaging a resolution test target, as shown in **Figure**

4.10, achieving a linewidth resolution of ~3.5 μm, roughly equivalent to the performance of a 4× objective lens with a numerical aperture (NA) of ~0.1. Compared to the presented system, which takes 87 s to scan an agar plate, a traditional lens-based bright-field microscope using a 4× objective lens would take approximately 128 min to scan a plate with the same diameter (60 mm), owing to the requirement for mechanical axial focusing (**Table 4.2**). In addition, the holographic imaging that is at the heart of this system provides better performance for early colony detection over bright-field imaging. Since bacteria can be considered phase objects, growth-related changes in a holographic image are enhanced compared to the bright-field images, enabling the earlier detection of bacterial growth and more sensitive measurements (**Figure 4.3**b).

Another important advantage of this system is the minimum requirement for optical alignment; the presented platform is tolerant towards structural changes, such as variations in the sample-to-sensor distance or the illumination angle. The computational refocusing capability also enables the screening of thick samples, e.g., melted agar plates. [225] An example of a 3D sample is illustrated in **Figure 4.11**, where *E. coli* colonies are formed at different depths inside the solid culture medium with a thickness of ~5 mm. For example, the colony marked with "A" grew at ~2170 μm measured from the surface of the agar, whereas the colony marked with "B" was on the agar surface. The presented system localizes colonies growing at different depths within a 3D culture medium using a single hologram measurement at each scanning position. However, it is a nontrivial task to image a 3D sample using a conventional lens-based microscope because of the time required for mechanical focusing and the refractive index mismatch between the culture medium and the air, which degrades the image resolution as a result of aberrations. Therefore, the corresponding bright-field microscopy images of the whole plates could only be acquired after 24 h of incubation.

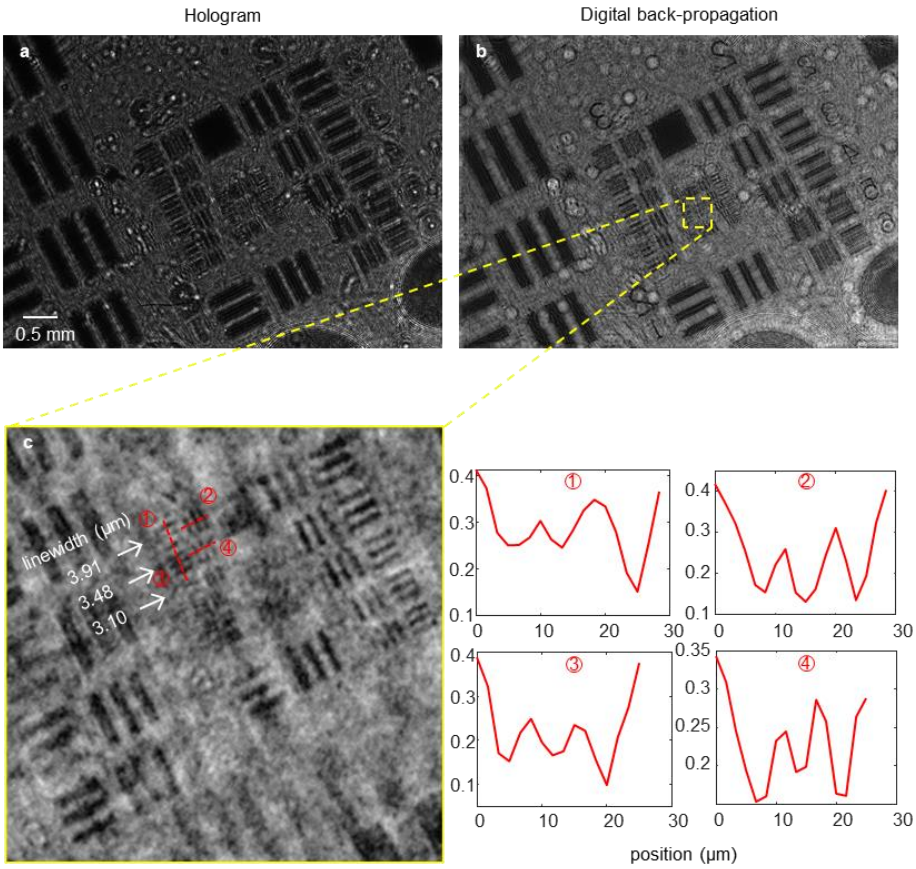

Figure 4.10 Resolution characterization of the lens-free bacterial colony detection system. (a) Raw hologram captured by the image sensor. (b) Digitally back-propagated hologram. (c) Zoomed-in region demonstrates a half-pitch resolution of ~3.5 μm.

The presented platform also employs a modular design that is scalable to a larger sample size and a smaller tile-scan time interval. The monitoring field of view (FOV) of this platform is fundamentally limited by the image acquisition time and the stage moving speed. With further optimization of the hardware and control algorithms, an imaging throughput of > 50 cm^2/min can be reached. Alternatively, several image sensors can be installed and connected to a single computer for high-throughput parallel imaging. [226] In this proof-of-concept implementation, the image processing for each time interval takes ~20 min and fits well into the 30 min measurement period between each scan. In case a shorter time interval is desired, an image processing procedure

implemented using MATLAB and Python/PyTorch programming environments can be further accelerated by programming in C/C++. With the help of graphic processing units (GPUs), one can expect >10-fold time savings in computation. [49]

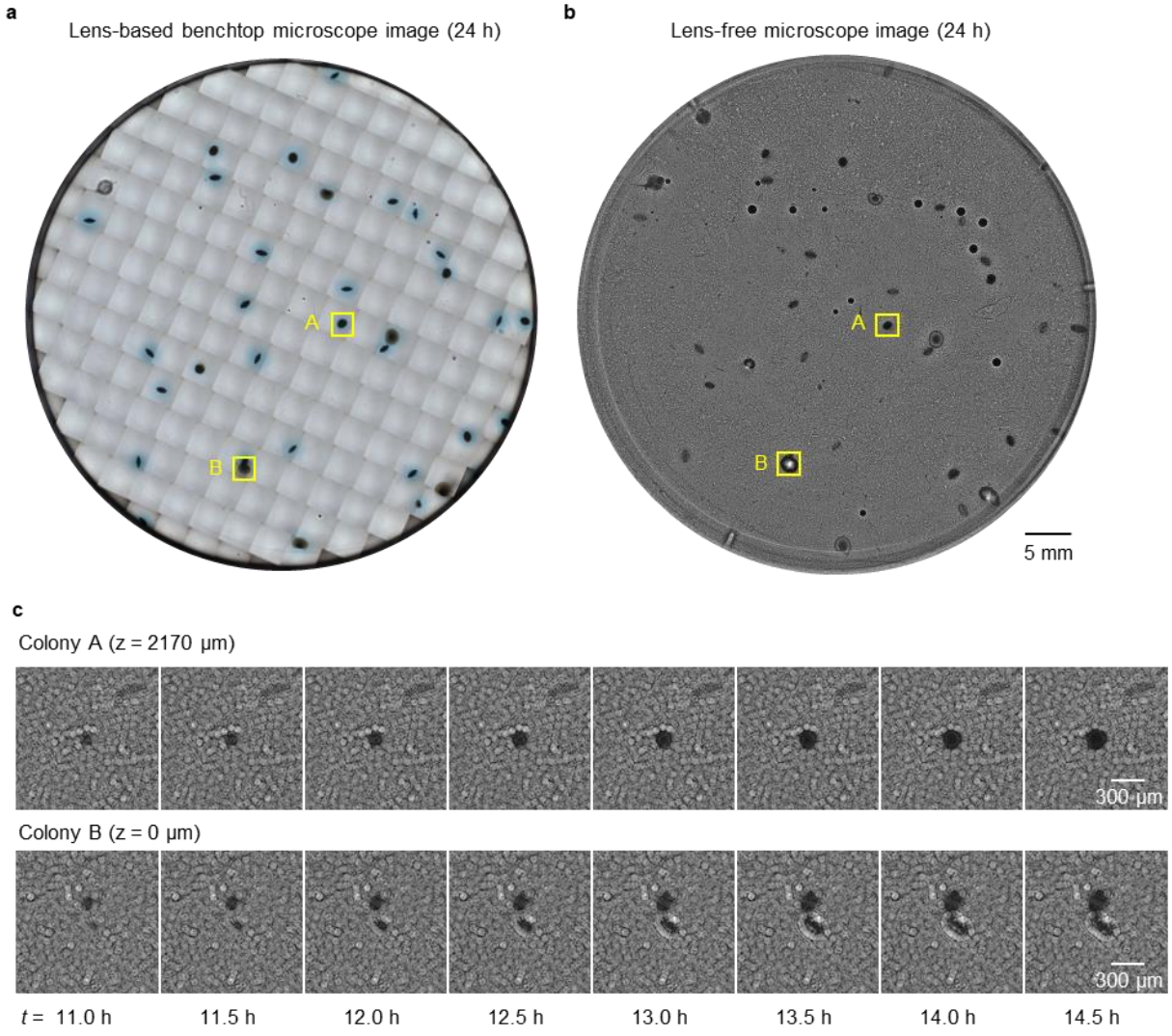

Figure 4.11 *E. coli* colonies grew at different depths within the 3D culture medium. (a) Image of the sample plate captured using a lens-based benchtop microscope after 24 hours of incubation and stitched by the microscope software. (b) Image of the sample plate captured using the lens-free microscope at 24 h of incubation. (c) Images of 2 colonies marked in (a) and (b) that grew at different depths, axially separated by ~2.17 mm.

This unique platform is integrated with an incubator to keep the agar plates at a desired temperature. The incubator is a thermal glass plate that contains uniform lines of optically clear indium tin oxide (ITO) electrode for heating the sample placed on top. This system is controlled with a controller, which is lightweight. Throughout the experiments, the temperature at the agar surface where bacteria grew at ~38 °C was set so that all of the tested bacterial species could grow and develop colonies. This temperature was not optimized to promote the growth of a specific species. Therefore, the adjustment of the incubation environment, temperature and humidity can potentially be used to further accelerate colony growth and help us achieve earlier detection and identification of specific bacterial colonies. Another important parameter for the growth of microorganisms is the humidity. This system can also be integrated with a controlled humidity chamber for better control and analysis of the growth dynamics of various microorganisms. [227]

In summary, I presented a deep learning-based live bacterium monitoring system for the early detection of growing colonies and the classification of colony species using deep learning. I demonstrated a proof-of-concept device using 3 types of bacteria, i.e., *E. coli*, *K. aerogenes*, and *K. pneumoniae*, and achieved > 12 h time savings for both the early detection and the classification of growing species compared to the gold-standard EPA-approved methods. Achieving an LOD of ~1 CFU/L in ≤ 9 h, I believe that this versatile system will not only benefit water and food quality monitoring but also provide a powerful tool for microbiology research.

Chapter 5 Conclusions

Computational microscopy has been an indispensable tool for many decades for it offers imaging capabilities beyond direct observation through a microscope. By combining the recent development of deep learning with computational imaging, a whole new area is opened up for more possibilities and technical revolutions. This dissertation starts with basic computational microscopy concepts and introduces an out-of-focus pixel super-resolution (OFI-PSR) technique based on traditional iterative optimization approach. This classical approach, although increases the throughput of a coherent microscopy system with minimum changes to the hardware, does not provide real-time performance and requires physical modeling of the image formation process. Then I introduced deep learning enhanced mobile phone microscopy as an example of the learning-based approaches that do not rely on prior knowledge of the imaging system and enables cross-modality image super-resolution and transformation. The concept of cross-modality image transformation, e.g., from mobile phone microscopic images to the equivalence from a benchtop microscope in this case, can be vastly expanded to more applications, which are then introduced in the following chapters.

In summary of Chapter 2, my contributions to deep learning-based single image super-resolution include: (1) I developed a framework of microscopic imaging, data processing, and network training for cross-modality image super-resolution. This framework trains a neural network based on pure experimental data therefore does not rely on any modeling or approximation of the physical system, and achieves image transformations even when analytically models cannot be built, e.g. TIRF to TIRF-SIM image transformation, confocal to STED microscopic image transformation. (2) I also introduced several methods to quantitatively evaluate

the network inference quality using, e.g., SSIM and SNR metrics, PSF characterization, artifact analysis.

Deep learning cross-modality image transformation can also be applied to fluorescence to bright-field microscopic image transformation, which introduces an exciting application of virtual histological staining of unlabeled tissue sections. In summary of Chapter 3, I contributed to this work as one of the leading researchers together with my colleagues and invented the framework of virtual histological staining of unlabeled tissue sections using auto-fluorescence images and deep learning. This is also a systematic framework that includes methods of autofluorescence imaging, image pre-processing, deep neural network model trainings, and network inference evaluation. The

Beyond image transformation, I have also explored using deep learning for object detection in high-dimensional data. In summary of Chapter 4, I demonstrated early detection and classification or bacterial colonies using deep learning and coherent imaging. Here I built a coherent imaging system that is both cost-effective and high-throughput, which can image a whole 60 mm-diameter agar plate in 87 s. Then I developed pseudo-3D deep neural networks to detect early stage colonies from 4-dimension (time-x-y-amplitude/phase) image stacks and classify their species. Using this framework, I achieved 90% detection rate within 7–10 h with a precision of 99.2-100%, and ~80% classification accuracy of all tested species within 12 h which represent a total time savings of more than 12 h compared to the gold-standard methods.

Deep learning has revolutionized the field of computational microscopy, achieved superior performance in computation speed and image quality in many areas, and brings a lot more opportunities that will introduce paradigm shifts in many areas. The deep learning-based computational techniques relieve hardware requirements and democratize complex and high-cost

imaging modalities to general users. Deep learning techniques also help reduce data dimensions and extract key information from complex non-intuitive dataset, therefore, greatly improve the efficiency of optical imaging and sensing systems. I believe deep learning computational imaging techniques will be the fundamental tools in further research and field applications.

References

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

2. H. Wang, Z. Göröcs, W. Luo, Y. Zhang, Y. Rivenson, L. A. Bentolila, and A. Ozcan, "Computational out-of-focus imaging increases the space–bandwidth product in lens-based coherent microscopy," Optica **3**, 1422–1429 (2016).

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

4. T. Liu, K. de Haan, B. Bai, Y. Rivenson, Y. Luo, H. Wang, D. Karalli, H. Fu, Y. Zhang, J. FitzGerald, and A. Ozcan, "Deep learning-based holographic polarization microscopy," ArXiv200700741 Phys. (2020).

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

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

7. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: Super Resolution Single Molecule Microscopy by Deep Learning," ArXiv180109631 Phys. (2018).

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

9. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nat. Biomed. Eng. **3**, 466–477 (2019).

10. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," Light Sci. Appl. **8**, 23 (2019).

11. Y. Zhang, K. de Haan, Y. Rivenson, J. Li, A. Delis, and A. Ozcan, "Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue," Light Sci. Appl. **9**, 78 (2020).

12. Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light Sci. Appl. **7**, e17141 (n.d.).

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

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

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

16. P. Ferraro, S. Grilli, D. Alfieri, S. De Nicola, A. Finizio, G. Pierattini, B. Javidi, G. Coppola, and V. Striano, "Extended focused image in microscopy by digital holography," Opt. Express **13**, 6738 (2005).

17. F. Charrière, A. Marian, F. Montfort, J. Kuehn, T. Colomb, E. Cuche, P. Marquet, and C. Depeursinge, "Cell refractive index tomography by digital holographic microscopy," Opt. Lett. **31**, 178 (2006).

18. F. Charrière, N. Pavillon, T. Colomb, C. Depeursinge, T. J. Heger, E. A. D. Mitchell, P. Marquet, and B. Rappaz, "Living specimen tomography by digital holographic microscopy: morphometry of testate amoeba," Opt. Express **14**, 7005 (2006).

19. L. Miccio, D. Alfieri, S. Grilli, P. Ferraro, A. Finizio, L. D. Petrocellis, and S. D. Nicola, "Direct full compensation of the aberrations in quantitative phase microscopy of thin objects by a single digital hologram," Appl. Phys. Lett. **90**, 041104 (2007).

20. B. Kemper and G. von Bally, "Digital holographic microscopy for live cell applications and technical inspection," Appl. Opt. **47**, A52 (2008).

21. V. Micó, Z. Zalevsky, C. Ferreira, and J. García, "Superresolution digital holographic microscopy for three-dimensional samples," Opt. Express **16**, 19260 (2008).

22. Y.-S. Choi and S.-J. Lee, "Three-dimensional volumetric measurement of red blood cell motion using digital holographic microscopy," Appl. Opt. **48**, 2983 (2009).

23. W. M. Ash III, L. Krzewina, and M. K. Kim, "Quantitative imaging of cellular adhesion by total internal reflection holographic microscopy," Appl. Opt. **48**, H144 (2009).

24. T. Tahara, K. Ito, T. Kakue, M. Fujii, Y. Shimozato, Y. Awatsuji, K. Nishio, S. Ura, T. Kubota, and O. Matoba, "Parallel phase-shifting digital holographic microscopy," Biomed. Opt. Express **1**, 610 (2010).

25. C. Fang-Yen, W. Choi, Y. Sung, C. J. Holbrow, R. R. Dasari, and M. S. Feld, "Video-rate tomographic phase microscopy," J. Biomed. Opt. **16**, 011005-011005–5 (2011).

26. P. Memmolo, G. Di Caprio, C. Distante, M. Paturzo, R. Puglisi, D. Balduzzi, A. Galli, G. Coppola, and P. Ferraro, "Identification of bovine sperm head for morphometry analysis in quantitative phase-contrast holographic microscopy," Opt. Express **19**, 23215 (2011).

27. J. Min, B. Yao, P. Gao, R. Guo, B. Ma, J. Zheng, M. Lei, S. Yan, D. Dan, T. Duan, Y. Yang, and T. Ye, "Dual-wavelength slightly off-axis digital holographic microscopy," Appl. Opt. **51**, 191 (2012).

28. A. Anand, V. K. Chhaniwal, N. R. Patel, and B. Javidi, "Automatic Identification of Malaria-Infected RBC With Digital Holographic Microscopy Using Correlation Algorithms," IEEE Photonics J. **4**, 1456–1464 (2012).

29. P. Gao, B. Yao, J. Min, R. Guo, B. Ma, J. Zheng, M. Lei, S. Yan, D. Dan, and T. Ye, "Autofocusing of digital holographic microscopy based on off-axis illuminations," Opt. Lett. **37**, 3630 (2012).

30. P. Petruck, R. Riesenberg, and R. Kowarschik, "Optimized coherence parameters for high-resolution holographic microscopy," Appl. Phys. B **106**, 339–348 (2011).

31. A. El Mallahi, C. Minetti, and F. Dubois, "Automated three-dimensional detection and classification of living organisms using digital holographic microscopy with partial spatial

coherent source: application to the monitoring of drinking water resources," Appl. Opt. **52**, A68 (2013).

32. P. Gao, G. Pedrini, and W. Osten, "Structured illumination for resolution enhancement and autofocusing in digital holographic microscopy," Opt. Lett. **38**, 1328 (2013).

33. J. Kostencka, T. Kozacki, and K. Liżewski, "Autofocusing method for tilted image plane detection in digital holographic microscopy," Opt. Commun. **297**, 20–26 (2013).

34. A. Anand, A. Faridian, V. K. Chhaniwal, S. Mahajan, V. Trivedi, S. K. Dubey, G. Pedrini, W. Osten, and B. Javidi, "Single beam Fourier transform digital holographic quantitative phase microscopy," Appl. Phys. Lett. **104**, 103705 (2014).

35. X. Yu, J. Hong, C. Liu, M. Cross, D. T. Haynie, and M. K. Kim, "Four-dimensional motility tracking of biological cells by digital holographic microscopy," J. Biomed. Opt. **19**, 045001–045001 (2014).

36. Y. Zhang, W. Jiang, L. Tian, L. Waller, and Q. Dai, "Self-learning based Fourier ptychographic microscopy," Opt. Express **23**, 18471 (2015).

37. B. Mandracchia, V. Pagliarulo, M. Paturzo, and P. Ferraro, "Surface Plasmon Resonance Imaging by Holographic Enhanced Mapping," Anal. Chem. **87**, 4124–4128 (2015).

38. S. Mahajan, V. Trivedi, P. Vora, V. Chhaniwal, B. Javidi, and A. Anand, "Highly stable digital holographic microscope using Sagnac interferometer," Opt. Lett. **40**, 3743 (2015).

39. N. Verrier, C. Fournier, and T. Fournel, "3D tracking the Brownian motion of colloidal particles using digital holographic microscopy and joint reconstruction," Appl. Opt. **54**, 4996 (2015).

40. F. Yi, I. Moon, and B. Javidi, "Cell morphology-based classification of red blood cells using holographic imaging informatics," Biomed. Opt. Express **7**, 2385 (2016).

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

42. "Microscope Digital Camara DP80," http://www.olympus-lifescience.com/en/camera/color/dp80/#!cms[tab]=%2Fcamera%2Fcolor%2Fdp80%2Fresources.

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

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

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

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

47. E. McLeod and A. Ozcan, "Unconventional methods of imaging: computational microscopy and compact implementations," Rep. Prog. Phys. **79**, 076001 (2016).

48. A. Ozcan and E. McLeod, "Lensless Imaging and Sensing," Annu. Rev. Biomed. Eng. **18**, 77–102 (2016).

49. S. O. Isikman, W. Bishara, S. Mavandadi, F. W. Yu, S. Feng, R. Lau, and A. Ozcan, "Lens-free optical tomographic microscope with a large imaging volume on a chip," Proc. Natl. Acad. Sci. **108**, 7296–7301 (2011).

50. W. Luo, Y. Zhang, A. Feizi, Z. Göröcs, and A. Ozcan, "Pixel super-resolution using wavelength scanning," Light Sci. Appl. **5**, e16060 (2015).

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

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

53. "Olympus IX83 Inverted Microscope," http://www.olympus-lifescience.com/en/microscopes/inverted/ix83/#!cms[tab]=%2Fmicroscopes%2Finverted%2Fix83%2Ffeaturesca50677bd9a5846f8deb5d96a828969.

54. C.-S. Guo, L. Zhang, H.-T. Wang, J. Liao, and Y. Y. Zhu, "Phase-shifting error and its elimination in phase-shifting digital holography," Opt. Lett. **27**, 1687 (2002).

55. P. Guo and A. J. Devaney, "Digital microscopy using phase-shifting digital holography with two reference waves," Opt. Lett. **29**, 857 (2004).

56. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," ArXiv150504597 Cs (2015).

57. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," ArXiv14062661 Cs Stat (2014).

58. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," ArXiv14126980 Cs (2014).

59. A. Ozcan, "Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools," Lab Chip **14**, 3187–3194 (2014).

60. D. N. Breslauer, R. N. Maamari, N. A. Switz, W. A. Lam, and D. A. Fletcher, "Mobile Phone Based Clinical Microscopy for Global Health Applications," PLoS ONE **4**, e6320 (2009).

61. Y. Lu, W. Shi, J. Qin, and B. Lin, "Low cost, portable detection of gold nanoparticle-labeled microfluidic immunoassay with camera cell phone," ELECTROPHORESIS **30**, 579–582 (2009).

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

63. H. Zhu, S. Mavandadi, A. F. Coskun, O. Yaglidere, and A. Ozcan, "Optofluidic Fluorescent Imaging Cytometry on a Cell Phone," Anal. Chem. **83**, 6641–6647 (2011).

64. Z. J. Smith, K. Chu, A. R. Espenson, M. Rahimzadeh, A. Gryshuk, M. Molinaro, D. M. Dwyre, S. Lane, D. Matthews, and S. Wachsmann-Hogiu, "Cell-Phone-Based Platform for Biomedical Device Development and Education Applications," PLOS ONE **6**, e17150 (2011).

65. V. Oncescu, D. O'Dell, and D. Erickson, "Smartphone based health accessory for colorimetric detection of biomarkers in sweat and saliva," Lab. Chip **13**, 3232–3238 (2013).

66. P. B. Lillehoj, M.-C. Huang, N. Truong, and C.-M. Ho, "Rapid electrochemical detection on a mobile phone," Lab. Chip **13**, 2950–2955 (2013).

67. H. C. Koydemir, Z. Gorocs, D. Tseng, B. Cortazar, S. Feng, R. Y. L. Chan, J. Burbano, E. McLeod, and A. Ozcan, "Rapid imaging, detection and quantification of Giardia lamblia cysts using mobile-phone based fluorescent microscopy and machine learning," Lab Chip **15**, 1284–1293 (2015).

68. S. Feng, D. Tseng, D. Di Carlo, O. B. Garner, and A. Ozcan, "High-throughput and automated diagnosis of antimicrobial resistance using a cost-effective cellphone-based micro-plate reader," Sci. Rep. **6**, (2016).

69. M. Kühnemund, Q. Wei, E. Darai, Y. Wang, I. Hernández-Neuta, Z. Yang, D. Tseng, A. Ahlford, L. Mathot, T. Sjöblom, A. Ozcan, and M. Nilsson, "Targeted DNA sequencing and in situ mutation analysis using mobile phone microscopy," Nat. Commun. **8**, 13913 (2017).

70. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

71. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in (2016), pp. 770–778.

72. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D.

Hassabis, "Mastering the game of Go with deep neural networks and tree search," Nature **529**, 484–489 (2016).

73. N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," Science **353**, 790–794 (2016).

74. V. N. Murthy, S. Maji, and R. Manmatha, "Automatic Image Annotation Using Deep Learning Representations," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, ICMR '15 (ACM, 2015), pp. 603–606.

75. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature **542**, 115–118 (2017).

76. D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," Annu. Rev. Biomed. Eng. **19**, 221–248 (2017).

77. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," (2015).

78. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," (2016).

79. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437–1443 (2017).

80. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep Learning for Photoacoustic Tomography from Sparse Data," (2017).

81. M. Mardani, E. Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, M. Alley, N. Thakur, S. Han, W. Dally, J. M. Pauly, and L. Xing, "Deep Generative Adversarial Networks for Compressed Sensing Automates MRI," (2017).

82. "Nokia Lumia 1020 Camera - Sensor and Lens Explained," http://www.cameradebate.com/2013/nokia-lumia-1020-camera-sensor-lens/.

83. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vis. **60**, 91–110 (2004).

84. "Correlation coefficients - MATLAB corrcoef," https://www.mathworks.com/help/matlab/ref/corrcoef.html.

85. S. Culley, D. Albrecht, C. Jacobs, P. M. Pereira, C. Leterrier, J. Mercer, and R. Henriques, "NanoJ-SQUIRREL: quantitative mapping and minimisation of super-resolution optical imaging artefacts," (2017).

86. "NanoJ-Core-ImageJ Plugin," https://bitbucket.org/rhenriqueslab/nanoj-core/wiki/Home.

87. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in (2016), pp. 1874–1883.

88. D. Han, J. Kim, and J. Kim, "Deep Pyramidal Residual Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 6307–6315.

89. A. Kingston, A. Sakellariou, T. Varslot, G. Myers, and A. Sheppard, "Reliable automatic alignment of tomographic projection data by passive auto-focus," Med. Phys. **38**, 4934–4945 (2011).

90. CIE 116-1995, "Industrial Colour-Difference Evaluation," (1995).

91. Y. Zhang, Y. Wu, Y. Zhang, and A. Ozcan, "Color calibration and fusion of lens-free and mobile-phone microscopy images for high-resolution and accurate color reproduction," Sci. Rep. **6**, srep27811 (2016).

92. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).

93. P. Hamel, M. E. P. Davies, K. Yoshii, and M. Goto, "Transfer Learning In MIR: Sharing Learned Latent Representations For Music Audio Classification And Similarity," (2013).

94. A. Badano, C. Revie, A. Casertano, W.-C. Cheng, P. Green, T. Kimpe, E. Krupinski, C. Sisson, S. Skrøvseth, D. Treanor, P. Boynton, D. Clunie, M. J. Flynn, T. Heki, S. Hewitt, H. Homma, A. Masia, T. Matsui, B. Nagy, M. Nishibori, J. Penczek, T. Schopf, Y. Yagi, and H. Yokoi, "Consistency and Standardization of Color in Medical Imaging: a Consensus Report," J. Digit. Imaging **28**, 41–52 (2015).

95. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution," Science **313**, 1642–1645 (2006).

96. S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy," Biophys. J. **91**, 4258–4272 (2006).

97. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nat. Methods **3**, 793–796 (2006).

98. S. van de Linde, A. Löschberger, T. Klein, M. Heidbreder, S. Wolter, M. Heilemann, and M. Sauer, "Direct stochastic optical reconstruction microscopy with standard fluorescent probes," Nat. Protoc. **6**, 991–1009 (2011).

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

100. M. G. L. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," J. Microsc. **198**, 82–87 (2000).

101. S. Cox, "Super-resolution imaging in live cells," Dev. Biol. **401**, 175–181 (2015).

102. M. G. L. Gustafsson, "Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution," Proc. Natl. Acad. Sci. **102**, 13081–13086 (2005).

103. R. Henriques, M. Lelek, E. F. Fornasiero, F. Valtorta, C. Zimmer, and M. M. Mhlanga, "QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ," Nat. Methods **7**, 339–340 (2010).

104. A. Small and S. Stahlheber, "Fluorophore localization algorithms for super-resolution microscopy," Nat. Methods **11**, 267–279 (2014).

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

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

107. S. Culley, D. Albrecht, C. Jacobs, P. M. Pereira, C. Leterrier, J. Mercer, and R. Henriques, "Quantitative mapping and minimization of super-resolution optical imaging artifacts," Nat. Methods **15**, 263–266 (2018).

108. D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, "Quantitative evaluation of software packages for single-molecule localization microscopy," Nat. Methods **12**, 717–724 (2015).

109. P. Almada, S. Culley, and R. Henriques, "PALM and STORM: Into large fields and high-throughput microscopy with sCMOS detectors," Methods **88**, 109–121 (2015).

110. T. Wilson and B. R. Masters, "Confocal microscopy," Appl. Opt. **33**, 565–566 (1994).

111. D. Li, L. Shao, B.-C. Chen, X. Zhang, M. Zhang, B. Moses, D. E. Milkie, J. R. Beach, J. A. Hammer, M. Pasham, T. Kirchhausen, M. A. Baird, M. W. Davidson, P. Xu, and E. Betzig, "Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics," Science **349**, aab3500 (2015).

112. W. H. Richardson, "Bayesian-Based Iterative Method of Image Restoration," J. Opt. Soc. Am. 1917-1983 **62**, 55 (1972).

113. L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron. J. **79**, 745 (1974).

114. L. Landweber, "An Iteration Formula for Fredholm Integral Equations of the First Kind," Am. J. Math. **73**, 615–624 (1951).

115. J. N. Farahani, M. J. Schibler, and L. A. Bentolila, "Stimulated emission depletion (STED) microscopy: from theory to practice," Microsc. Sci. Technol. Appl. Educ. **2**, 1539–1547 (2010).

116. P. Hamel, M. E. P. Davies, K. Yoshii, and M. Goto, "Transfer Learning In MIR: Sharing Learned Latent Representations For Music Audio Classification And Similarity," https://ai.google/research/pubs/pub41530.

117. S. Wäldchen, J. Lehmann, T. Klein, S. van de Linde, and M. Sauer, "Light-induced cell damage in live-cell super-resolution microscopy," Sci. Rep. **5**, 15348 (2015).

118. B. Hein, K. I. Willig, and S. W. Hell, "Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell," Proc. Natl. Acad. Sci. **105**, 14271–14276 (2008).

119. B. Hein, K. I. Willig, C. A. Wurm, V. Westphal, S. Jakobs, and S. W. Hell, "Stimulated Emission Depletion Nanoscopy of Living Cells Using SNAP-Tag Fusion Proteins," Biophys. J. **98**, 158–163 (2010).

120. M. Dyba and S. W. Hell, "Photostability of a fluorescent marker under pulsed excited-state depletion through stimulated emission," Appl. Opt. **42**, 5123–5129 (2003).

121. P. Kner, B. B. Chhun, E. R. Griffis, L. Winoto, and M. G. L. Gustafsson, "Super-resolution video microscopy of live cells by structured illumination," Nat. Methods **6**, 339–342 (2009).

122. D. Leyton-Puig, T. Isogai, E. Argenzio, B. van den Broek, J. Klarenbeek, H. Janssen, K. Jalink, and M. Innocenti, "Flat clathrin lattices are dynamic actin-controlled hubs for clathrin-mediated endocytosis and signalling of specific receptors," Nat. Commun. **8**, 16068 (2017).

123. R. Fiolka, L. Shao, E. H. Rego, M. W. Davidson, and M. G. L. Gustafsson, "Time-lapse two-color 3D imaging of live cells with doubled resolution using structured illumination," Proc. Natl. Acad. Sci. **109**, 5311–5315 (2012).

124. J. P. Ferguson, N. M. Willy, S. P. Heidotting, S. D. Huber, M. J. Webber, and C. Kural, "Deciphering dynamics of clathrin-mediated endocytosis in a living organism," J. Cell Biol. **214**, 347–358 (2016).

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

126. R. Liu and J. Jia, "Reducing boundary artifacts in image deconvolution," in *2008 15th IEEE International Conference on Image Processing* (2008), pp. 505–508.

127. S. Culley, D. Albrecht, C. Jacobs, P. M. Pereira, C. Leterrier, J. Mercer, and R. Henriques, "Quantitative mapping and minimization of super-resolution optical imaging artifacts," Nat. Methods **15**, 263–266 (2018).

128. L. A. Bentolila, R. Prakash, D. Mihic-Probst, M. Wadehra, H. K. Kleinman, T. S. Carmichael, B. Péault, R. L. Barnhill, and C. Lugassy, "Imaging of Angiotropism/Vascular Co-Option in a Murine Model of Brain Melanoma: Implications for Melanoma Progression along Extravascular Pathways," Sci. Rep. **6**, 23834 (2016).

129. F. Aguet, S. Upadhyayula, R. Gaudin, Y. Chou, E. Cocucci, K. He, B.-C. Chen, K. Mosaliganti, M. Pasham, W. Skillern, W. R. Legant, T.-L. Liu, G. Findlay, E. Marino, G. Danuser, S. Megason, E. Betzig, T. Kirchhausen, and J. Lippincott-Schwartz, "Membrane dynamics of dividing cells imaged by lattice light-sheet microscopy," Mol. Biol. Cell **27**, 3418–3435 (2016).

130. N. M. Willy, J. P. Ferguson, S. D. Huber, S. P. Heidotting, E. Aygün, S. A. Wurm, E. Johnston-Halperin, M. G. Poirier, and C. Kural, "Membrane mechanics govern spatiotemporal heterogeneity of endocytic clathrin coat dynamics," Mol. Biol. Cell **28**, 3480–3488 (2017).

131. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," Nat. Methods **9**, 676–682 (2012).

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

133. D. Sage, D. Prodanov, J.-Y. Tinevez, and J. Schindelin, "MIJ: Making interoperability between ImageJ and Matlab possible," in *ImageJ User & Developer Conference* (2012).

134. Y. Rivenson, H. Wang, Z. Wei, Y. Zhang, H. Gunaydin, and A. Ozcan, "Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue," ArXiv180311293 Phys. (2018).

135. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).

136. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," ArXiv14126980 Cs (2014).

137. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," ArXiv160508695 Cs (2016).

138. F. Aguet, D. V. D. Ville, and M. Unser, "Model-Based 2.5-D Deconvolution for Extended Depth of Field in Brightfield Microscopy," IEEE Trans. Image Process. **17**, 1144–1153 (2008).

139. M. Born, E. Wolf, and A. B. Bhatia, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. (Cambridge University Press, 1999).

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

141. D. Sage, L. Donati, F. Soulez, D. Fortun, G. Schmit, A. Seitz, R. Guiet, C. Vonesch, and M. Unser, "DeconvolutionLab2: An open-source software for deconvolution microscopy," Methods **115**, 28–41 (2017).

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

143. Y. Katznelson, "An Introduction to Harmonic Analysis, Dover Publications, New York, 1976," (n.d.).

144. Y. K. Tao, D. Shen, Y. Sheikine, O. O. Ahsen, H. H. Wang, D. B. Schmolze, N. B. Johnson, J. S. Brooker, A. E. Cable, J. L. Connolly, and J. G. Fujimoto, "Assessment of breast pathologies using nonlinear microscopy," Proc. Natl. Acad. Sci. **111**, 15304–15309 (2014).

145. S. Witte, A. Negrean, J. C. Lodder, C. P. J. de Kock, G. Testa Silva, H. D. Mansvelder, and M. Louise Groot, "Label-free live brain imaging and targeted patching with third-harmonic generation microscopy," Proc. Natl. Acad. Sci. U. S. A. **108**, 5970–5975 (2011).

146. M. Ji, D. A. Orringer, C. W. Freudiger, S. Ramkissoon, X. Liu, D. Lau, A. J. Golby, I. Norton, M. Hayashi, N. Y. R. Agar, G. S. Young, C. Spino, S. Santagata, S. Camelo-Piragua, K. L. Ligon, O. Sagher, and X. S. Xie, "Rapid, label-free detection of brain tumors with stimulated Raman scattering microscopy," Sci. Transl. Med. **5**, 201ra119 (2013).

147. F.-K. Lu, S. Basu, V. Igras, M. P. Hoang, M. Ji, D. Fu, G. R. Holtom, V. A. Neel, C. W. Freudiger, D. E. Fisher, and X. S. Xie, "Label-free DNA imaging in vivo with stimulated Raman scattering microscopy," Proc. Natl. Acad. Sci. U. S. A. **112**, 11624–11629 (2015).

148. D. A. Orringer, B. Pandian, Y. S. Niknafs, T. C. Hollon, J. Boyle, S. Lewis, M. Garrard, S. L. Hervey-Jumper, H. J. L. Garton, C. O. Maher, J. A. Heth, O. Sagher, D. A. Wilkinson, M. Snuderl, S. Venneti, S. H. Ramkissoon, K. A. McFadden, A. Fisher-Hubbard, A. P. Lieberman, T. D. Johnson, X. S. Xie, J. K. Trautman, C. W. Freudiger, and S. Camelo-Piragua, "Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy," Nat. Biomed. Eng. **1**, 0027 (2017).

149. H. Tu, Y. Liu, D. Turchinovich, M. Marjanovic, J. K. Lyngsø, J. Lægsgaard, E. J. Chaney, Y. Zhao, S. You, W. L. Wilson, B. Xu, M. Dantus, and S. A. Boppart, "Stain-free histopathology by programmable supercontinuum pulses," Nat. Photonics **10**, 534–540 (2016).

150. F. Fereidouni, Z. T. Harmany, M. Tian, A. Todd, J. A. Kintner, J. D. McPherson, A. D. Borowsky, J. Bishop, M. Lechpammer, S. G. Demos, and R. Levenson, "Microscopy with ultraviolet surface excitation for rapid slide-free histology," Nat. Biomed. Eng. **1**, 957–966 (2017).

151. A. K. Glaser, N. P. Reder, Y. Chen, E. F. McCarty, C. Yin, L. Wei, Y. Wang, L. D. True, and J. T. C. Liu, "Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens," Nat. Biomed. Eng. **1**, 0084 (2017).

152. F. Jamme, S. Kascakova, S. Villette, F. Allouche, S. Pallu, V. Rouam, and M. Réfrégiers, "Deep UV autofluorescence microscopy for cell biology and tissue histology," Biol. Cell **105**, 277–288 (2013).

153. M. Monici, "Cell and tissue autofluorescence research and diagnostic applications," in *Biotechnology Annual Review* (Elsevier, 2005), Vol. 11, pp. 227–256.

154. A. C. Croce and G. Bottiroli, "Autofluorescence Spectroscopy and Imaging: A Tool for Biomedical Research and Diagnosis," Eur. J. Histochem. EJH **58**, (2014).

155. Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting Cancer Metastases on Gigapixel Pathology Images," ArXiv170302442 Cs (2017).

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

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

158. "Price List Effective June 1, 2017 | Histology Laboratory | Pathology Research Resources at Miller School of Medicine," http://cpl.med.miami.edu/pathology-research/histology-laboratory/price-list.

159. "Fee Schedule | Pathology & Laboratory Medicine," https://pathology.weill.cornell.edu/research/translational-research-services/fee-schedule.

160. "Research Histology - Rates | UC Davis Health System Department of Pathology," http://www.ucdmc.ucdavis.edu/pathology/research/research_labs/histology/rates.html.

161. I. A. Cree, Z. Deans, M. J. L. Ligtenberg, N. Normanno, A. Edsjö, E. Rouleau, F. Solé, E. Thunnissen, W. Timens, E. Schuuring, E. Dequeker, S. Murray, M. Dietel, P. Groenen, and J. H. Van Krieken, "Guidance for laboratories performing molecular pathology for cancer patients," J. Clin. Pathol. **67**, 923–931 (2014).

162. P. G. Patel, S. Selvarajah, S. Boursalie, N. E. How, J. Ejdelman, K.-P. Guerard, J. M. Bartlett, J. Lapointe, P. C. Park, J. B. A. Okello, and D. M. Berman, "Preparation of Formalin-fixed Paraffin-embedded Tissue Cores for both RNA and DNA Extraction," JoVE J. Vis. Exp. e54299–e54299 (2016).

163. H. Cho, S. Lim, G. Choi, and H. Min, "Neural Stain-Style Transfer Learning using GAN for Histopathological Images," ArXiv171008543 Cs (2017).

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

165. P. H. S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," Comput. Vis. Image Underst. **78**, 138–156 (2000).

166. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. (Cambridge University Press, 2003).

167. Y. Rivenson, H. C. Koydemir, H. Wang, Z. Wei, Z. Ren, H. Gunaydin, Y. Zhang, Z. Gorocs, K. Liang, D. Tseng, and A. Ozcan, "Deep learning enhanced mobile-phone microscopy," ArXiv171204139 Phys. (2017).

168. D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context Encoders: Feature Learning by Inpainting," ArXiv E-Prints **1604**, arXiv:1604.07379 (2016).

169. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in (IEEE, 2017), pp. 5967–5976.

170. Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Gunaydin, Y. Zhang, Z. Gorocs, K. Liang, D. Tseng, and A. Ozcan, "Deep learning enhanced mobile-phone microscopy," ACS Photonics (2018).

171. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," ArXiv150504597 Cs (2015).

172. K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds., Lecture Notes in Computer Science (Springer International Publishing, 2016), pp. 630–645.

173. "Convert RGB color values to YCbCr color space - MATLAB rgb2ycbcr," https://www.mathworks.com/help/images/ref/rgb2ycbcr.html.

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

175. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," Nat. Methods **9**, 676–682 (2012).

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

177. "Zoomify—Zoomable web images!," http://zoomify.com/.

178. "GIGAmacro: Exploring Small Things in a Big Way," https://viewer.gigamacro.com/.

179. B. J. Vakoc, R. M. Lanning, J. A. Tyrrell, T. P. Padera, L. A. Bartlett, T. Stylianopoulos, L. L. Munn, G. J. Tearney, D. Fukumura, R. K. Jain, and B. E. Bouma, "Three-dimensional microscopy of the tumor microenvironment *in vivo* using optical frequency domain imaging," Nat. Med. **15**, 1219–1223 (2009).

180. S. Kozikowski, L. Wolfram, and R. Alfano, "Fluorescence spectroscopy of eumelanins," IEEE J. Quantum Electron. **20**, 1379–1382 (1984).

181. M. Elleder and J. Borovanský, "Autofluorescence of melanins induced by ultraviolet radiation and near ultraviolet light. A histochemical and biochemical study," Histochem. J. **33**, 273–281 (2001).

182. R. D. Lovchik, G. V. Kaigala, M. Georgiadis, and E. Delamarche, "Micro-immunohistochemistry using a microfluidic probe," Lab. Chip **12**, 1040–1043 (2012).

183. A. Sandgren, M. Strong, P. Muthukrishnan, B. K. Weiner, G. M. Church, and M. B. Murray, "Tuberculosis drug resistance mutation database," PLoS Med. **6**, e1000002 (2009).

184. T. M. Arain, A. E. Resconi, M. J. Hickey, and C. K. Stover, "Bioluminescence screening in vitro (Bio-Siv) assays for high-volume antimycobacterial drug discovery.," Antimicrob. Agents Chemother. **40**, 1536–1541 (1996).

185. W. R. Jacobs, R. G. Barletta, R. Udani, J. Chan, G. Kalkut, G. Sosne, T. Kieser, G. J. Sarkis, G. F. Hatfull, and B. R. Bloom, "Rapid assessment of drug susceptibilities of Mycobacterium tuberculosis by means of luciferase reporter phages," Science **260**, 819–822 (1993).

186. R. Goodacre, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell, and P. J. Rooney, "Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks," Microbiology **144**, 1157–1170 (1998).

187. J.-C. Lagier, G. Dubourg, M. Million, F. Cadoret, M. Bilen, F. Fenollar, A. Levasseur, J.-M. Rolain, P.-E. Fournier, and D. Raoult, "Culturing the human microbiota and culturomics," Nat. Rev. Microbiol. 1 (2018).

188. N. Fierer, C. L. Lauber, N. Zhou, D. McDonald, E. K. Costello, and R. Knight, "Forensic identification using skin bacterial communities," Proc. Natl. Acad. Sci. **107**, 6477–6481 (2010).

189. H. C. Koydemir, Z. Gorocs, D. Tseng, B. Cortazar, S. Feng, R. Y. L. Chan, J. Burbano, E. McLeod, and A. Ozcan, "Rapid imaging, detection and quantification of Giardia lamblia cysts using mobile-phone based fluorescent microscopy and machine learning," Lab. Chip **15**, 1284–1293 (2015).

190. S. P. Oliver, B. M. Jayarao, and R. A. Almeida, "Foodborne pathogens in milk and the dairy farm environment: food safety and public health implications," Foodbourne Pathog. Dis. **2**, 115–129 (2005).

191. "World Water Day," (n.d.).

192. S. DeFlorio-Barker, C. Wing, R. M. Jones, and S. Dorevitch, "Estimate of incidence and cost of recreational waterborne illness on United States surface waters," Environ. Health **17**, 3 (2018).

193. United States, ed., *Method 1604: Total Coliforms and Escherichia Coli in Water by Membrane Filtration Using a Simultaneous Detection Technique (MI Medium)* (United States, Environmental Protection Agency, Office of Water, 2002).

194. "Current Waterborne Disease Burden Data & Gaps | Healthy Water | CDC," https://www.cdc.gov/healthywater/burden/current-data.html.

195. US EPA, "Analytical Methods Approved for Compliance Monitoring under the Long Term 2 Enhanced Surface Water Treatment Rule," (2017).

196. R. A. Deshmukh, K. Joshi, S. Bhand, and U. Roy, "Recent developments in detection and enumeration of waterborne bacteria: a retrospective minireview," MicrobiologyOpen **5**, 901–922 (2016).

197. R. Amann and B. M. Fuchs, "Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques," Nat. Rev. Microbiol. **6**, 339–348 (2008).

198. D.-K. Kang, M. M. Ali, K. Zhang, S. S. Huang, E. Peterson, M. A. Digman, E. Gratton, and W. Zhao, "Rapid detection of single bacteria in unprocessed blood using Integrated Comprehensive Droplet Digital Detection," Nat. Commun. **5**, 5427 (2014).

199. *Title 40: Protection of Environment* (n.d.), Vol. 136.3.

200. K. Huff, A. Aroonnual, A. E. F. Littlejohn, B. Rajwa, E. Bae, P. P. Banada, V. Patsekin, E. D. Hirleman, J. P. Robinson, G. P. Richards, and A. K. Bhunia, "Light-scattering sensor for real-time identification of Vibrio parahaemolyticus, Vibrio vulnificus and Vibrio cholerae colonies on solid agar plate," Microb. Biotechnol. **5**, 607–620 (2012).

201. J. Choi, J. Yoo, M. Lee, E.-G. Kim, J. S. Lee, S. Lee, S. Joo, S. H. Song, E.-C. Kim, J. C. Lee, H. C. Kim, Y.-G. Jung, and S. Kwon, "A rapid antimicrobial susceptibility test based on single-cell morphological analysis," Sci. Transl. Med. **6**, 267ra174-267ra174 (2014).

202. Y. Jo, S. Park, J. Jung, J. Yoon, H. Joo, M. Kim, S.-J. Kang, M. C. Choi, S. Y. Lee, and Y. Park, "Holographic deep learning for rapid optical screening of anthrax spores," Sci. Adv. **3**, e1700606 (2017).

203. S. O. Van Poucke and H. J. Nelis, "A 210-min solid phase cytometry test for the enumeration of Escherichia coli in drinking water," J. Appl. Microbiol. **89**, 390–396 (2000).

204. M. Kim, M. Pan, Y. Gai, S. Pang, C. Han, C. Yang, and S. K. Tang, "Optofluidic ultrahigh-throughput detection of fluorescent drops," Lab. Chip **15**, 1417–1423 (2015).

205. I. Tryland, H. Braathen, A. Wennberg, F. Eregno, and A.-L. Beschorner, "Monitoring of β-D-Galactosidase activity as a surrogate parameter for rapid detection of sewage contamination in urban recreational water," Water **8**, 65 (2016).

206. S. O. Van Poucke and H. J. Nelis, "Limitations of highly sensitive enzymatic presence-absence tests for detection of waterborne coliforms and Escherichia coli.," Appl Env. Microbiol **63**, 771–774 (1997).

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

208. *EPA Microbiological Alternate Test Procedure (ATP) Protocol for Drinking Water, Ambient Water, Wastewater, and Sewage Sludge Monitoring Methods* (United States, Environmental Protection Agency, Office of Water, 2010).

209. "CHROMagar™ ECC Product Leaflet," (n.d.).

210. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," ArXiv160806993 Cs (2016).

211. J. A. Shapiro, "The significances of bacterial colony patterns," BioEssays **17**, 597–607 (1995).

212. P.-T. Su, C.-T. Liao, J.-R. Roan, S.-H. Wang, A. Chiou, and W.-J. Syu, "Bacterial Colony from Two-Dimensional Division to Three-Dimensional Development," PLOS ONE **7**, e48098 (2012).

213. F. D. Farrell, M. Gralka, O. Hallatschek, and B. Waclaw, "Mechanical interactions in bacterial colonies and the surfing probability of beneficial mutations," J. R. Soc. Interface **14**, (2017).

214. Sheats Julian, Sclavi Bianca, Cosentino Lagomarsino Marco, Cicuta Pietro, and Dorfman Kevin D., "Role of growth rate on the orientational alignment of Escherichia coli in a slit," R. Soc. Open Sci. **4**, 170463 (n.d.).

215. M. W. LeChevallier and G. A. McFeters, "Enumerating Injured Coliforms in Drinking Water," J. Am. Water Works Assoc. **77**, 81–87 (1985).

216. "CDC-Salmonella-Factsheet," https://www.cdc.gov/salmonella/pdf/CDC-Salmonella-Factsheet.pdf.

217. H. Liu, C. A. Whitehouse, and B. Li, "Presence and Persistence of Salmonella in Water: The Impact on Microbial Quality of Water and Food Safety," Front. Public Health **6**, (2018).

218. J. R. Hutchison, M. W. Widder, S. M. Brooks, L. M. Brennan, L. Souris, V. T. Divito, W. H. van der Schalie, and R. M. Ozanich, "Consistent production of chlorine-stressed bacteria from non-chlorinated secondary sewage effluents for use in the U.S. Environmental Protection Agency Alternate Test Procedure protocol," J. Microbiol. Methods **163**, 105651 (2019).

219. "Colilert 18 - IDEXX US," https://www.idexx.com/en/water/water-products-services/colilert-18/.

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

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

222. Z. Qiu, T. Yao, and T. Mei, "Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks," ArXiv171110305 Cs (2017).

223. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization.," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015).

224. "Alternate Test Procedures in Clean Water Act Analytical Methods," (n.d.).

225. E. R. Sanders, "Aseptic Laboratory Techniques: Plating Methods," J. Vis. Exp. JoVE (2012).

226. Y. Zhang, H. C. Koydemir, M. M. Shimogawa, S. Yalcin, A. Guziak, T. Liu, I. Oguz, Y. Huang, B. Bai, Y. Luo, Y. Luo, Z. Wei, H. Wang, V. Bianco, B. Zhang, R. Nadkarni, K. Hill, and A. Ozcan, "Motility-based label-free detection of parasites in bodily fluids using holographic speckle analysis and deep learning," Light Sci. Appl. **7**, 108 (2018).

227. M. P. Cobo, S. Libro, N. Marechal, D. D'Entremont, D. P. Cobo, and M. Berkmen, "Visualizing bacterial colony morphologies using time-lapse imaging chamber MOCHA," J. Bacteriol. **200**, e00413-17 (2018).