UNIVERSITY OF CALIFORNIA

Los Angeles

Computational deep learning microscopy

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical and Computer Engineering

by

Kevin de Haan

2022

ABSTRACT OF THE DISSERTATION

Computational deep learning microscopy

by

Kevin de Haan

Doctor of Philosophy in Electrical & Computer Engineering

University of California, Los Angeles, 2022

Professor Aydogan Ozcan, Chair

Over the past decade, deep learning has become one of the leading techniques used in the field of image processing. Beyond popular tasks in computer vision such as classification and segmentation, it has proven to be revolutionary in its applications for image enhancement and transformations. It has significantly changed the field of computational optics – and neural networks can now be used to accurately and rapidly solve a wide variety of inverse problems in microscopy.

This dissertation discusses a few major classes of inverse imaging problems that can be solved using deep learning. The dissertation first presents, a framework that can be used to enhance microscopy images using single image super-resolution . This framework has been proven to be effective at super-resolving images captured with a holographic microscope that are resolution limited both by the number of pixels used for imaging, as well as by the numerical aperture (NA) of the microscope. The effectiveness of this same general framework beyond optical microscopy, will be further demonstrated by super resolving electron microscopy images.

Next, the dissertation will show that a similar super-resolution framework can be extended to perform a transformation between two imaging modalities and improve the overall quality of images by using it to enhance images of thin blood smears captured by a cost-effective mobile-phone microscope. By enhancing mobile phone microscopy images to match the quality of a top-of-the-line benchtop microscope, the images are standardized, have their resolution improved and have aberrations removed, allowing the images to be used for screening of sickle cell disease. Using a deep learning based classification framework, 98% accuracy was achieved during blind tested of 96 human blood smear slides. Furthermore, by enhancing the images, the image quality is brought to a level which can be used by clinicians for further analysis if required.

Finally, this same framework will be used to transform microscopy images and generate images from that are equivalent to those which have undergone chemical labeling and show some of the many applications of this technology. The technique was applied to virtual staining of label-free thin histological tissue sections which were used to generate multiple stains from a single tissue section, enabling different stains to be performed at the microscopic level, as well as blending of stains together – creating entirely new digital stains. This dissertation shows how multiple virtual stains can be used to generate synthetic datasets of perfectly matched stains, allowing downstream networks be trained to perform transformations between stains. The efficacy of three of these stain transformation networks – generating the Masson's trichrome, Jones silver stain, and periodic acid-Schiff stains from hematoxylin and eosin-stained kidney tissue are demonstrated in a diagnostic study, with the results showing the improvement that such technology can bring to patient care.

The dissertation of Kevin de Haan is approved.

Dino Di Carlo

Mona Jarrahi

Robert N. Candler

Aydogan Ozcan, Committee Chair

University of California, Los Angeles

2022

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to start by thanking my advisor, Professor Aydogan Ozcan for supporting me along my PhD journey. His guidance has been essential to my professional development, and he has helped me to grow as a researcher. I would also like to thank my doctoral committee for their advice on improving my research and in the preparation of this dissertation.

I would also like to acknowledge my lab colleagues for their valuable contributions to my research projects as well as their suggestions and guidance through my PhD. In particular, I would like to thank Dr. Yair Rivenson for his constant guidance, as well as other collaborators such as Tairan Liu, Zhensong Wei, Yijie Zhang, Hongda Wang, Çağatay Işıl, Zoltán Göröcs, and Hatice Ceylan Koydemir amongst others for all of the help and insight they provided. I have had such a great time working with so many talented scientists.

Vita

Kevin de Haan received his Bachelor of Science in Engineering Physics from Queens University in 2015. He then moved on to complete his Masters of Applied Science with Professor Stewart Aitchison at the University of Toronto in the Department of Electrical and Computer Engineering. In 2017 he joined Professor Aydogan Ozcan's Bio and Nanophotonics Lab in the Department of Electrical and Computer Engineering at UCLA. At UCLA his research has focused on performing image to image transformations to enhance the capability of microscope imaging systems.

Selected publications

1. **K. de Haan**, Y. Zhang, J.E. Zuckerman, T. Liu, A.E. Sisk, M.F.P. Diaz, K. Jen, A. Nobori, S. Liou, S. Zhang, R. Riahi, Y. Rivenson, W.D. Wallace, and A. Ozcan, "Deep learning-based transformation of H&E stained tissues into special stains," Nature Communications DOI: 10.1038/s41467-021-25221-2 (2021)
2. 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," ACS Photonics DOI: 10.1021/acsphotonics.0c01051 (2020)
3. 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: Science & Applications (Nature Publishing Group) DOI: 10.1038/s41377-020-0315-y (2020).
4. **K. de Haan**, H.C. Koydemir, Y. Rivenson, D. Tseng, E. Van Dyne, L.S. Bakic, D. Karinca, K. Liang, M. Ilango, E. Gumustekin, and A. Ozcan, "Automated screening of sickle cells using a smartphone-based microscope and deep learning," npj Digital Medicine DOI: 10.1038/s41746- 020-0282-y (2020).
5. **K. de Haan**, Y. Rivenson, Y. Wu, and A. Ozcan, "Deep learning-based image reconstruction and enhancement in optical microscopy," Proceedings of the IEEE DOI: 0.1109/JPROC.2019.2949575 (2019).
6. **K. de Haan**, Z.S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, "Resolution enhancement in scanning electron microscopy using deep learning," Scientific Reports DOI: 10.1038/s41598- 019-48444-2 (2019).
7. T. Liu, **K. de Haan**, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," Scientific Reports DOI: 10.1038/s41598- 019-40554-1 (2019).

8. Y. Rivenson, H. Wang, Z. Wei, **K. de Haan**, Y. Zhang, Y. Wu, H. Günaydın, J.E. Zuckerman, T. Chong, A.E. Sisk, L. M. Westbrook, W.D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering DOI: 10.1038/s41551-019-0362-y (2019).

Chapter 1 Solving inverse problems in microscopy using deep learning

Deep learning is a machine learning technique that involves the use of many layers of neural networks to perform automatic computation on a set of data. Deep learning has been used for a variety of computer vision tasks, such as image classification [1–3] and segmentation [4]. While neural networks were first proposed in the 1940s [5], and began to be used for processing documents in the 80s and 90s [6], there has been a significant revolution in both the quality and quantity of their use over the past decade. These improvements have been enabled by both software and hardware improvements, allowing deep learning to be used for many tasks for which it had not previously been feasible.

Deep learning has also been used to solve classical inverse problems in optics (i.e. super resolution [7,8] and image reconstruction [9]), and has gone beyond to perform tasks such as image-to-image translation [10]. In this dissertation I will demonstrate several applications of how deep learning can be used to solve both classical inverse problems in microscopy, as well as those which, while not traditionally considered inverse problems, can be constructed as one. Deep neural networks have a unique potential to solve these problems, as they do not rely upon an analytical model made by a user, but rather a set of data which can be experimentally generated. Therefore, many normally intractable problems can be solved, and models which rely upon simplifying assumptions can be improved upon.

In order to train a deep neural network to perform an image-to-image translation task, a large set of data is required. In this dissertation, I will mainly focus upon tasks where this data consists of a set of images which are used as the input and ground truth for the neural networks. Typically this data consists of matched, pairs of images, allowing the network to learn a direct transformation

between the pair of images (this is known as supervised training). However, in some cases synthetic or unmatched data can be used as well (the latter is often referred to as unsupervised training).

A significant amount of research continues to be performed to improve the performance of neural networks by changing architectures and other parameters. The majority of the networks shown in this dissertation are conditional generative adversarial networks (GANs) [11] and use variants of the U-net architecture as the generator [4]. As each application of the neural networks is different, the exact implementation of these networks will vary. Furthermore, tuning of parameters such as the number of layers and the number of channels in each layer in the network is necessary to ensure that the network is the correct size for a given dataset.

In this dissertation, I will apply these techniques to a broad spectrum of microscopy modalities. These include brightfield microscope, where white light is used to illuminate the sample, and after passing through the microscope is collected by an objective lens. I will also discuss digital holographic microscopy, which typically is performed by passing coherent light through a sample to create an interference pattern, which can be processed to reconstruct the image and extract the amplitude and phase of the light passing through the sample. Third, I will use scanning electron microscopy, which measures the intensity of an electron beam as a raster scan of the sample is performed. Electron microscopy is able to achieve much higher resolutions than optical microscopy techniques due to the low wavelength of the electrons within the beam. Finally, I will use fluorescence microscopy, which uses high energy photons to excite molecules within a given sample. These molecules re-emit the light at a higher wavelength (Stokes shift), which can be separated from the lower wavelength illuminating source by a filter, and measured.

Chapter 2 Super resolution for coherent imaging systems using deep learning

Super resolution has been one of the most commonly studied classical inverse problem for microscopy [12]. There are two main ways that microscopy images can be limited in resolution. They can be either pixel size-limited, where the pixel sampling frequency is the limiting factor for the resolution, or diffraction-limited, where the resolution is limited by the numerical aperture (NA) of the system. While deep learning and computer vision literature is concentrated on pixel-limited super-resolution, both are highly relevant in the field of microscopy. In this chapter, I present a deep learning-based framework that performs single-image super-resolution, focusing on coherent imaging systems. I will demonstrate that this framework can enhance the resolution of images captured by both pixel size-limited and diffraction-limited coherent imaging systems. The capabilities of this approach are experimentally validated by first super-resolving complex-valued images acquired using a lensfree on-chip holographic microscope, the resolution of which was pixel size-limited. I will also show that the same framework can be used to increase the resolution of a lens-based holographic imaging system that was limited in resolution by the numerical aperture of its objective lens. While in this chapter, the focus is on coherent imaging systems, similar frameworks can be applied to other microscope modalities such as scanning electron microscopes (discussed in Chapter 3) and brightfield microscopy (discussed in Chapter 4).

Parts of this chapter have been published in:

- T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, “Deep learning-based super-resolution in coherent imaging systems,” Scientific Reports DOI: 10.1038/s41598- 019-40554-1 (2019).

2.1 Introduction

Coherent imaging systems have many advantages for applications where the specimen's complex field information is of interest [14]. Since Gabor's seminal work, various optical and numerical techniques have been suggested [15] to acquire the complex field of a coherently illuminated specimen. This has allowed for the characterization of absorption and scattering properties of a sample, as well as enabling numerical refocusing at different depths within that sample volume. In order to infer an object's complex field in a coherent optical imaging system, the "missing phase" needs be retrieved. A classical solution to this missing phase problem is given by off-axis holography [16,17], which in general results in a reduction of the space-bandwidth product of the imaging system. In-line holographic imaging, which can be used to design compact microscopes [18], has utilized measurement diversity to generate a set of physical constraints for iterative phase retrieval [19–23]. Recently, deep-learning based holographic image reconstruction techniques have also been demonstrated to create a high-fidelity reconstruction from a single in-line hologram [9,24,25], and are capable of further extending the depth-of-field of the reconstructed image [26].

Several approaches have been demonstrated to improve the resolution of coherent imaging systems [27–32]. Most of these techniques require sequential measurements and assume that the object is quasi-static while a diverse set of measurements are performed on it. These measurements often require the use of additional hardware or sacrifice a degree of freedom such as the sample field-of-view [33]. In recent years, sparsity-based holographic reconstruction methods have also demonstrated that they are capable of increasing the resolution of coherent imaging systems without the need for additional measurements or hardware [34–37]. Sparse signal recovery

methods employed in coherent imaging are based on iterative optimization algorithms. These methods usually involve a comprehensive search over a parameter space to obtain the optimal object image and generally result in longer reconstruction times.

Deep learning-based approaches for super-resolution of incoherent microscopy modalities such as brightfield and fluorescence microscopy have also recently emerged [8,38–41]. These data-driven super-resolution approaches produce a trained deep convolutional neural network that learns to transform low-resolution images into high-resolution images in a single feed-forward (i.e., non-iterative) step. Generative adversarial networks (GANs) [11] are a form of deep neural network training framework that can be used to ensure that the generated image is sharp and realistic. A GAN is made up of two separate networks. A generator network is used to generate an image that has the same features as the label (ground truth) image, and a discriminator network tries to distinguish between the generated and label (ground truth) images.

In this chapter I apply deep learning to enhance the resolution of coherent imaging systems and demonstrate a conditional GAN that is trained to super-resolve both *pixel-limited* and *diffraction-limited* images. Furthermore, the success of this framework is demonstrated on biomedical samples such as thin sections of lung tissue and Papanicolaou (Pap) smear samples. These results are also quantified using the structural similarity index (SSIM) [42] and spatial frequency content of the network's output images in comparison to the higher resolution images (which constitute the ground truth). This data-driven image super-resolution framework is applicable to enhance the performance of various coherent imaging systems.

2.2 Methods

Here, I will briefly summarize the methods that are presented in this chapter; sub-sequent subsections will provide more information on specific methods employed in the work. The presented deep learning-based super-resolution approach was applied to two separate in-line holographic imaging geometries to demonstrate the efficacy of the technique. As illustrated in **Figure 2.1**a and **Figure 2.1**b, the two implemented configurations were a pixel size-limited system (to be referred to as System A) and a diffraction-limited coherent microscopy system (to be referred to as System B). Despite using different methods to create the super-resolved images, as a result of the different image formation models, both of these systems followed similar general hologram reconstruction steps, i.e., 1. Raw holograms were collected at different sample to sensor distances, 2. Autofocus was used to determine the accurate sample to sensor distances, 3. Phase was recovered using a multi-height phase recovery algorithm. These steps will be detailed in the following subsections of the Methods [18,20,43–46].

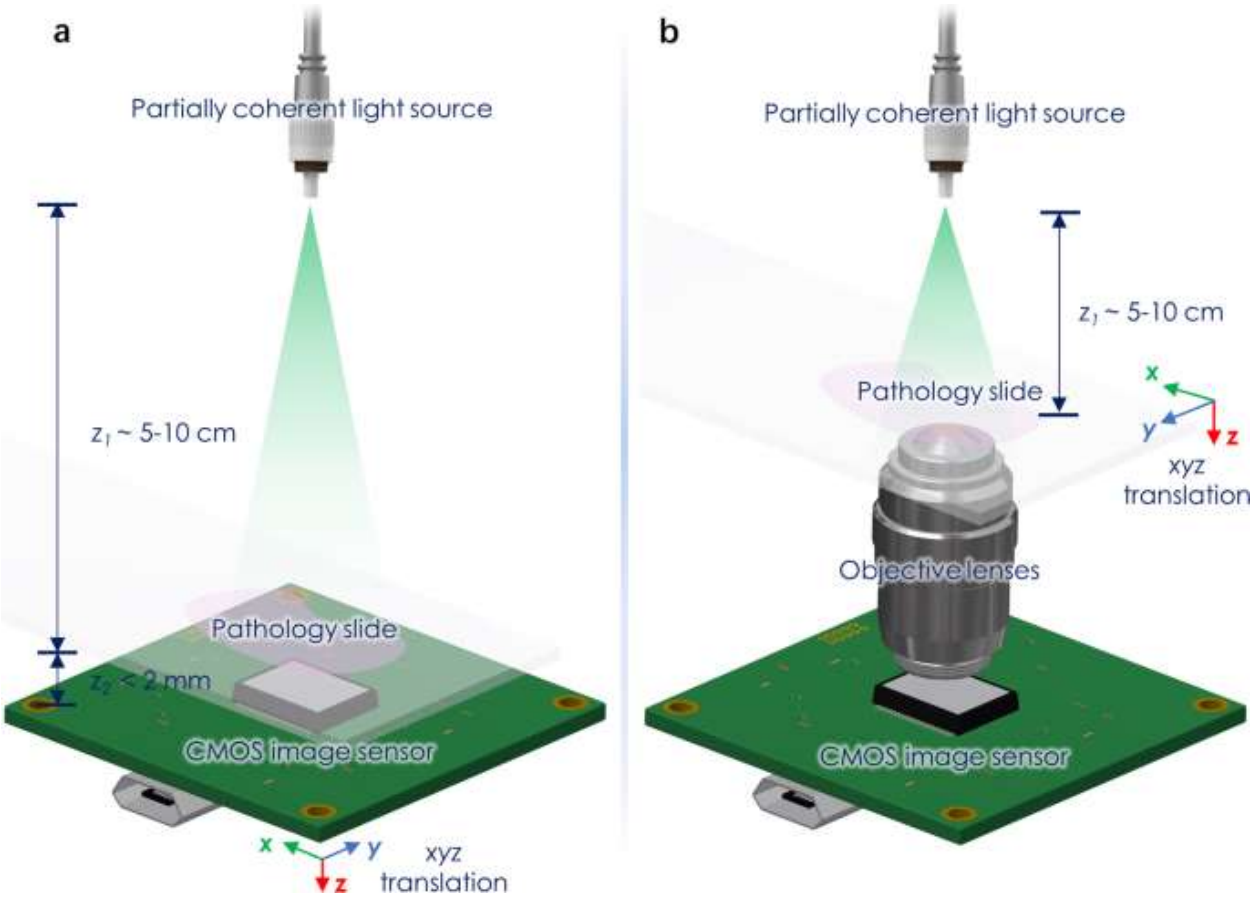

Figure 2.1 Schematic of the coherent imaging systems. (a) A Lens-free on-chip holographic microscope. The sample is placed at a short distance ($z_2 < 2$ mm) above the image sensor chip. The resolution of this lensless on-chip imaging modality (without the use of additional degrees of freedom) is pixel size-limited due to its unit magnification. (b) A lens-based in-line holographic microscope, implemented by removing the condenser and switching the illumination to a partially-coherent light source on a conventional bright-field microscope. The resolution in this case is limited by the NA of the objective lens.

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

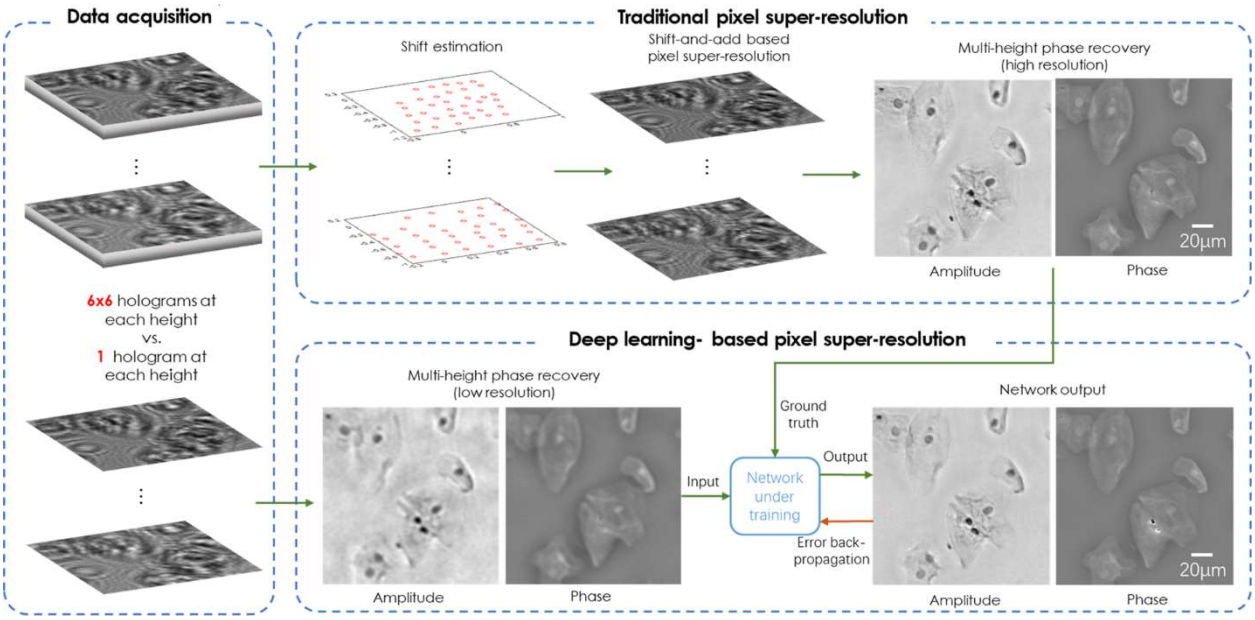

Figure 2.2 Schematic of the training process for deep-learning based pixel super-resolution. An outline of the data required to generate the network input and ground truth images is shown, together with an overview of how both the traditional super-resolution is performed and how the deep learning super-resolution network is trained.

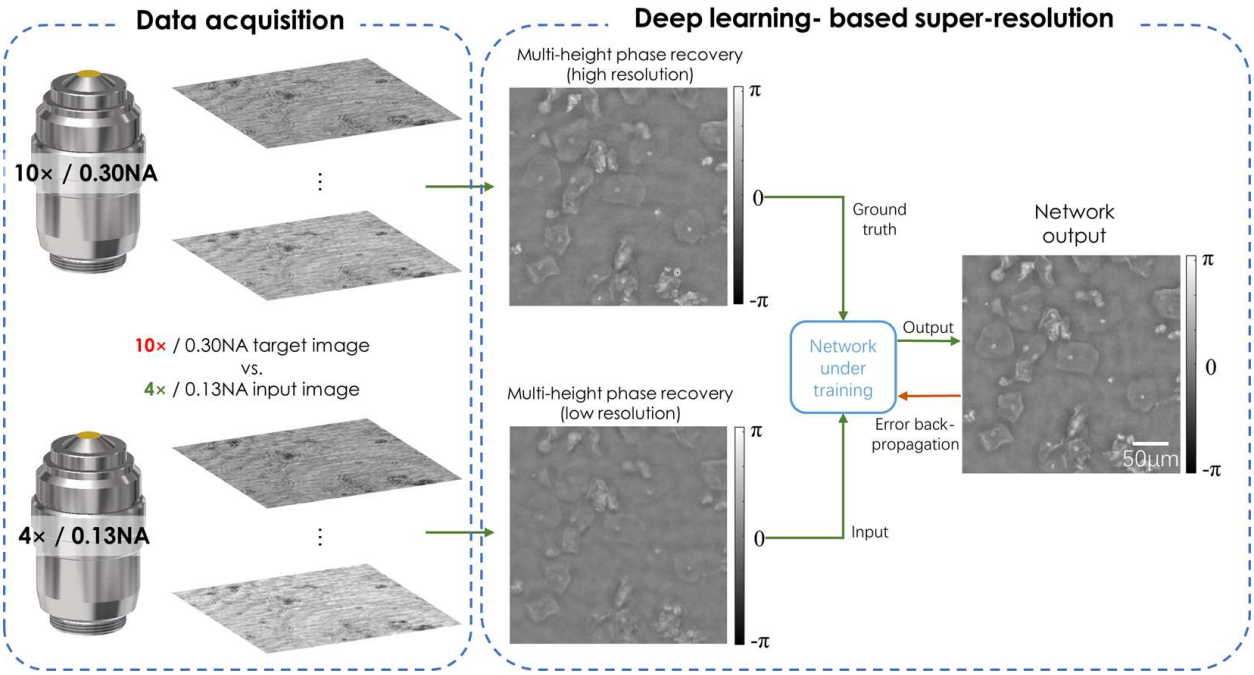

Figure 2.3 Schematic of the training process for deep learning-based optical super-resolution for an NA-limited coherent imaging system. An outline of the data required to generate the network input and ground truth images is shown, together with an overview of how the deep learning super-resolution network is trained.

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

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

For the pixel size-limited coherent imaging system (System A), the super-resolved images were created by collecting multiple low-resolution holograms at different lateral positions, where the CMOS image sensor was sub-pixel shifted by a mechanical stage to create a shift table. Once an accurate shift table was estimated, a shift-and-add based pixel super-resolution algorithm [43] was applied. The set-up used an illumination wavelength of 550 nm with a bandwidth ($\Delta\lambda$) of ~2 nm (WhiteLase Micro with acousto-optic tunable filter, NKT Photonics), a single mode fiber (QPMJ-3S2.5A-488-3.5/125-1-0.3-1) with a core diameter of ~3.5 μm and a source-to-sample distance (z_1) of ~5 cm. As a result, the effective spatial coherence diameter at the sensor plane was larger than the width of the CMOS imager chip used in the on-chip imaging system. Therefore, the achievable resolution is limited by the temporal coherence length of the illumination [47], which is defined as:

$$\Delta L_c \approx \sqrt{\frac{2\,ln\,2}{\pi}} \cdot \frac{\lambda^2}{n\Delta\lambda} = 100.47\ \mu m \tag{2.1}$$

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

$$NA = n\,sin\,\theta = n\sqrt{1 - cos^2\,\theta} = n\sqrt{1 - \left(\frac{z_2}{z_2 + \Delta L_c}\right)^2} \approx 0.6624 \tag{2.2}$$

Based on this effective numerical aperture and ignoring the pixel size at the hologram plane, the achievable coherence-limited resolution of the on-chip microscope is approximated as [17]:

$$d \propto \frac{\lambda}{NA} = \frac{0.55}{0.6624} = 0.8303\ \mu\text{m} \tag{2.3}$$

At the hologram/detector plane, however, the effective pixel pitch of the CMOS image sensor (IMX 081, Sony RGB sensor, pixel size of 1.12 μm) using only one color channel is 2.24 μm. Based on this, the effective pixel size for each super-resolved image after the application of the pixel super-resolution algorithm to 4 raw holograms (2×2 lateral positions), 9 raw holograms (3×3 lateral positions), and 36 raw holograms (6×6 lateral positions) are 1.12 μm, 0.7467 μm and 0.3733 μm, respectively. Based on Equation (2.3), the effective pixel size achieved by pixel super-resolution using 6×6 lateral positions can adequately sample the specimen's holographic diffraction pattern and is limited by temporal coherence. All of the other images (using 1×1, 2×2 and 3×3 raw holograms) remain pixel-limited in their achievable spatial resolution. This pixel-limited resolution of an on-chip holographic microscope is a result of its unit magnification. This allows the imaging system to have a large imaging field-of-view (FOV) that is only limited by the active area of the opto-electronic image sensor chip. This can easily reach 20-30 mm^2 and >10 cm^2 using state-of-the-art CMOS and CCD imagers, respectively [18].

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

size of ~1.625 μm. A 10×/0.30NA objective lens was used to acquire the higher resolution images (ground truth labels), achieving a resolution of 1.773 μm and an effective pixel size of ~0.65 μm.

Autofocusing and singular value decomposition-based background subtraction.

For both types of coherent imaging systems, holograms at 8 different sample-to-sensor distances were collected to perform the multi-height phase recovery [18,20,43–46]. This algorithm requires accurate knowledge of the sample-to-sensor distances used. These were estimated using an autofocusing algorithm. This algorithm assigned zero phase to the raw holograms collected by the image sensor and propagated them to different sample to sensor heights using the free space angular spectrum approach [17]. The Tamura of the gradient (ToG) edge sparsity-based criterion was computed [48] for each hologram and used to calculate the corresponding refocusing distance.

For the lens-based diffraction-limited coherent imaging system (System B), the autofocusing algorithm required an additional background subtraction step. For undesired particles or dust associated with the objective lens or other parts of the optical microscope, the diffraction pattern that is formed is independent of the sample and its position. Using this information, a singular value decomposition (SVD)-based background subtraction was performed [49], after which the ToG-based autofocusing algorithm was successfully applied.

Multi-height phase recovery.

The iterative multi-height phase recovery technique [44] was applied to eliminate the holographic image artifacts (twin image and self-interference terms [17]) in both of the coherent imaging systems that were used in this chapter. To perform this, an initial zero-phase was assigned to the intensity/amplitude measurement at the 1st hologram height. Next, the iterative algorithm begins by propagating the complex field to each hologram height until the 8^{th} height is reached,

and then backpropagates the resulting fields until the 1st height is reached. While the phase was retained at each hologram height, the amplitude was updated by averaging current amplitude and the square root of the measured intensity at each height.

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

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

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

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

score is used to determine which portion of the fused image is cropped out and is used as the input for the network.

GAN architecture and training process.

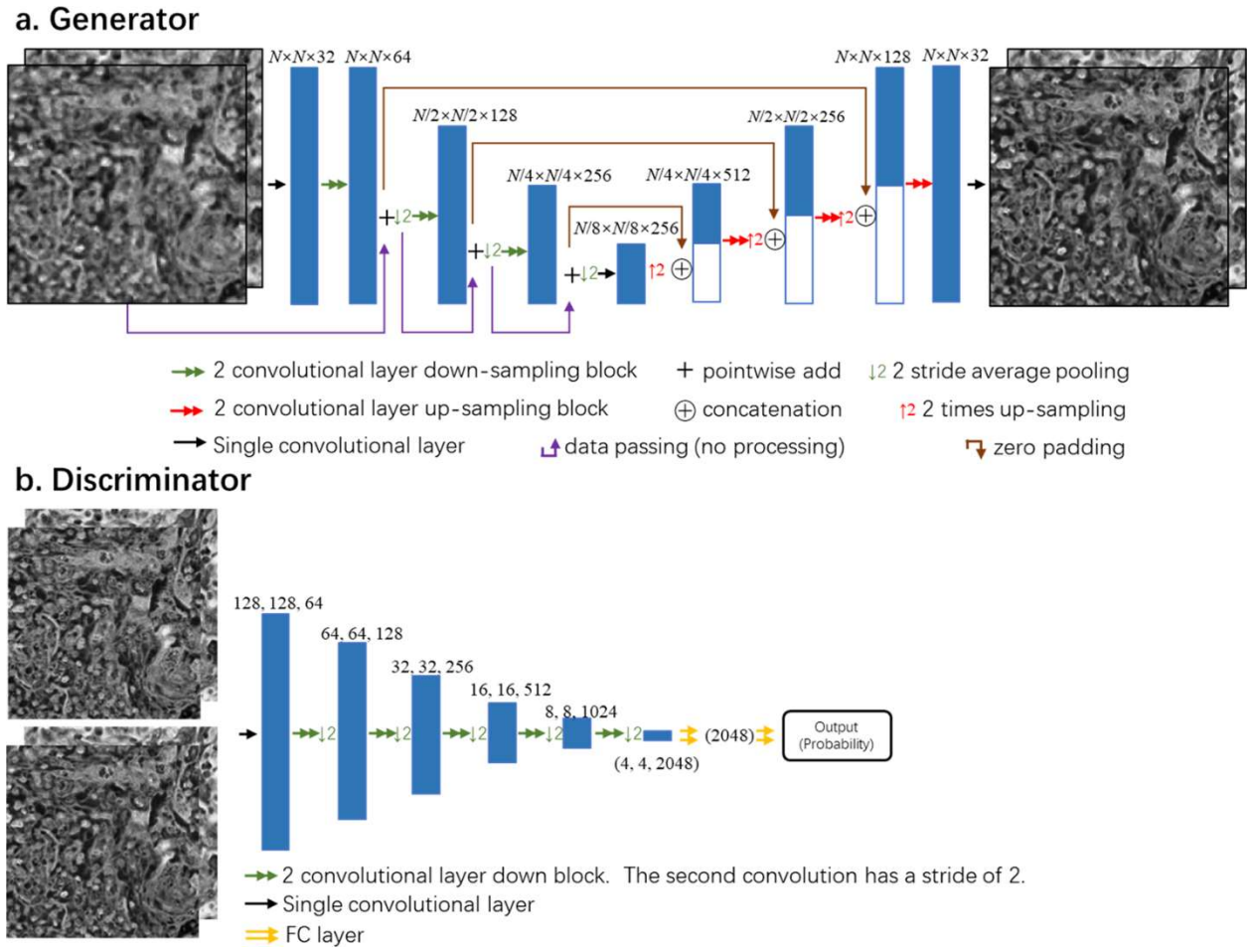

Figure 2.4 Diagram of the GAN structure. (a) Structure of the generator portion of the network. (b) Structure of the discriminator portion of the network.

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

Figure 2.4. For both the pixel-size limited and the diffraction-limited coherent imaging systems, the discriminator loss function is defined as:

$$l_{discriminator} = D\left(G\left(x_{input}\right)\right)^2 + \left(1 - D(z_{label})\right)^2 \tag{2.4}$$

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

$$l_{generator} = L_1\{z_{label}, G\left(x_{input}\right)\} + \gamma \times TV\{G\left(x_{input}\right)\} + \alpha \times \left(1 - D\left(G\left(x_{input}\right)\right)\right)^2 \tag{2.5}$$

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

$$L_1\{z_{\text{label}}, G\left(x_{\text{input}}\right)\} = \mathrm{E}_{\text{n\_pixels}}\left(E_{\text{n\_channels}}\left(\left|G\left(x_{\text{input}}\right) - z_{\text{label}}\right|\right)\right) \tag{2.6}$$

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

$$TV = E_{\text{n\_channels}}\left(\sum_{i,j}\left|G\left(x_{\text{input}}\right)_{i+1,j} - G\left(x_{\text{input}}\right)_{i,j}\right| + \left|G\left(x_{\text{input}}\right)_{i,j+1} - G\left(x_{\text{input}}\right)_{i,j}\right|\right) \tag{2.7}$$

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

The last term in Equation (2.5) (i.e., $\alpha\times(1\text{-}D(\mathrm{G}(x_{\mathrm{input}})))^2$) is a function of how well the output image of the generator network can be predicted by the discriminator network. α and γ are regularization parameters which were set to 0.00275 and 0.015 respectively. As a result of these parameters, the L_1 loss term, $L_1\{z_{\mathrm{label}},G(x_{\mathrm{input}})\}$, made up 60% of the overall loss, while the total variation term, $\gamma\times TV\{G(x_{\mathrm{input}})\}$, was approximately 0.25% of the total loss. The discriminator loss term, $\alpha\times(1\text{-}D(G(x_{\mathrm{input}})))^2$, made up the remainder of the overall generator loss. Once the networks were successfully trained, they reach a state of equilibrium where the discriminator network cannot successfully discriminate between the output and label images, and $D(G(x_{\mathrm{input}}))$ converged to approximately 0.5.

The loss function for the lens-based coherent microscope images (System B) incorporated an additional structural similarity index (SSIM) [42] term in addition to the terms included for the lensfree on-chip imaging system, i.e.,:

$$\begin{aligned} l_{\mathrm{generator}} = {} & L_1\left\{z_{\mathrm{label}}, G\left(x_{\mathrm{input}}\right)\right\} + \gamma\times TV\left\{G\left(x_{\mathrm{input}}\right)\right\} + \alpha\times\left(1-D\left(G\left(x_{\mathrm{input}}\right)\right)\right)^2 \\ & +\beta\times\mathrm{SSIM}\left\{G\left(x_{\mathrm{input}}\right), z_{\mathrm{label}}\right\} \end{aligned} \tag{2.8}$$

Where β is a regularization parameter set to 0.01, and $\mathrm{SSIM}\{x, z\}$ is defined as [42]:

$$\mathrm{SSIM}(x,z) = \frac{(2\mu_x\mu_z+c_1)(2\sigma_{x,z}+c_2)}{(\mu_x^2+\mu_z^2+c_1)(\sigma_x^2+\sigma_z^2+c_2)} \tag{2.9}$$

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

The generator network used an adapted U-net architecture [4]. The network began with a convolutional layer that increased the number of channels to 32 and a leaky rectified linear (LReLU) unit, defined as:

$$\mathrm{LReLU}(x) = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \quad (1)$$

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

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

The discriminator portion of the network was made up of a convolutional layer, followed by five discriminator blocks, an average pooling layer and two fully connected layers which reduced

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

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

Table 2.1 Training details for the deep neural networks. All the networks were trained with a batch size of 10 using 128×128 pixel patches.

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

Table 2.2 Time for each network to output a 1940×1940 pixel image. Each measurement is the average time, calculated using 150 test images.

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

Software implementation details.

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

Sample preparation.

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

2.3 Super-resolution of a pixel size-limited coherent imaging system (System A).

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

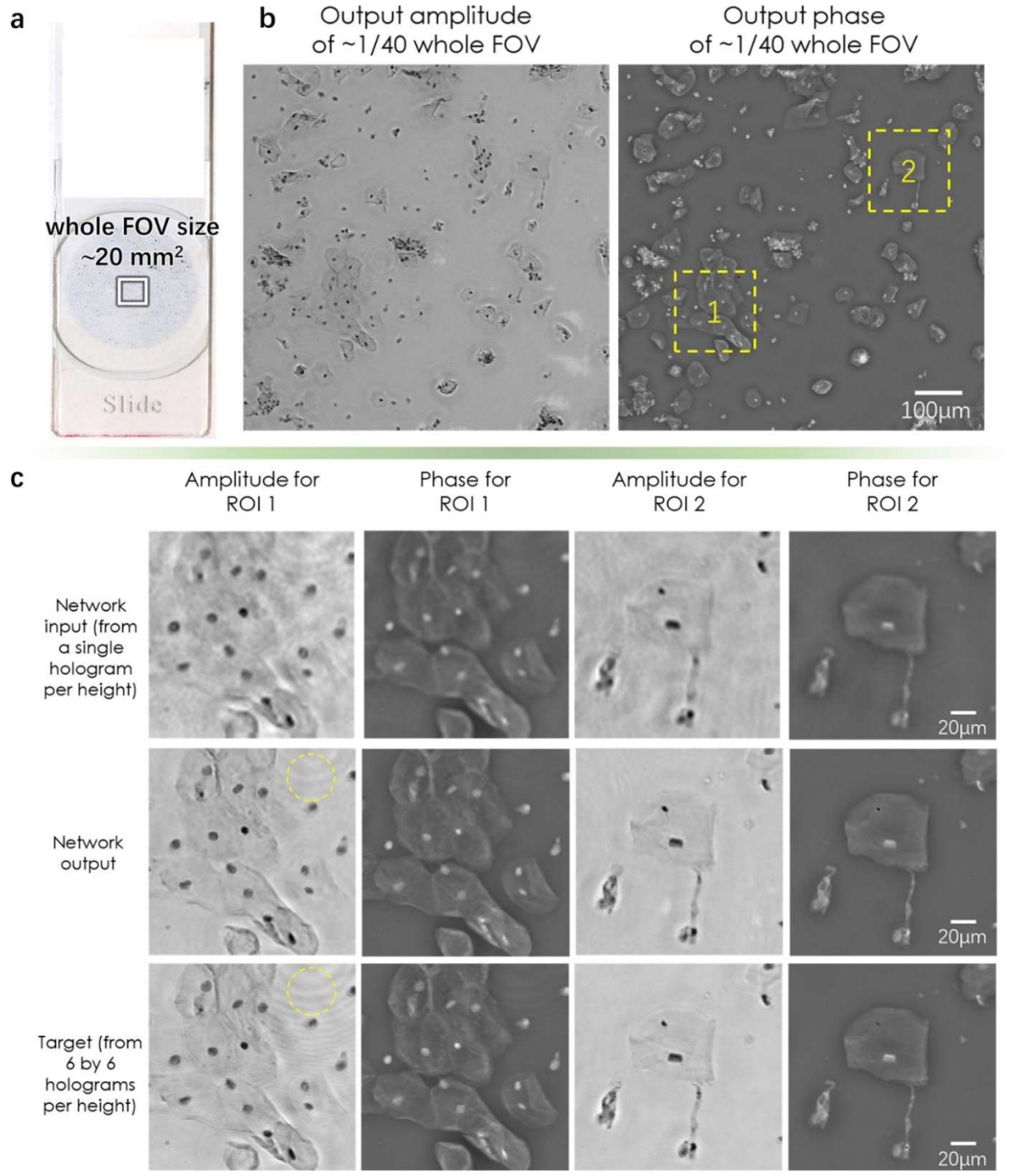

Figure 2.5 Visualized result for the pixel size-limited system. Deep learning-based pixel super-resolution imaging of a Pap smear slide under 550 nm illumination. (a) Whole FOV of the lensfree imaging system. (b) Amplitude and phase channels of the network output. (c) Further zoom-in of (b) for two regions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

Figure 2.5 illustrates the network's super-resolved output images along with pixel-size limited lower resolution input images and the higher resolution ground truth images of a Pap smear sample. The input images have a pixel pitch of 2.24 μm, and the label images have an effective pixel size of 0.37 μm (see the Methods section). For lung tissue sections, the efficacy of the super-resolution technique (**2.6**) was proven using three different deep networks, where the input images for each network used a different pixel size (2.24 μm, 1.12 μm, and 0.7467 μm, corresponding to 1×1, 2×2 and 3×3 lateral shifts, respectively, as detailed in the Methods section). Unlike the less densely connected Pap smear sample results, the network output is missing some of the spatial details that are seen by the high-resolution images of the lung tissue imaging when the input pixel size is at its coarsest level (2.24 μm pixel size). These spatial features are recovered by the two networks that use smaller input pixels as shown in **2.6**.

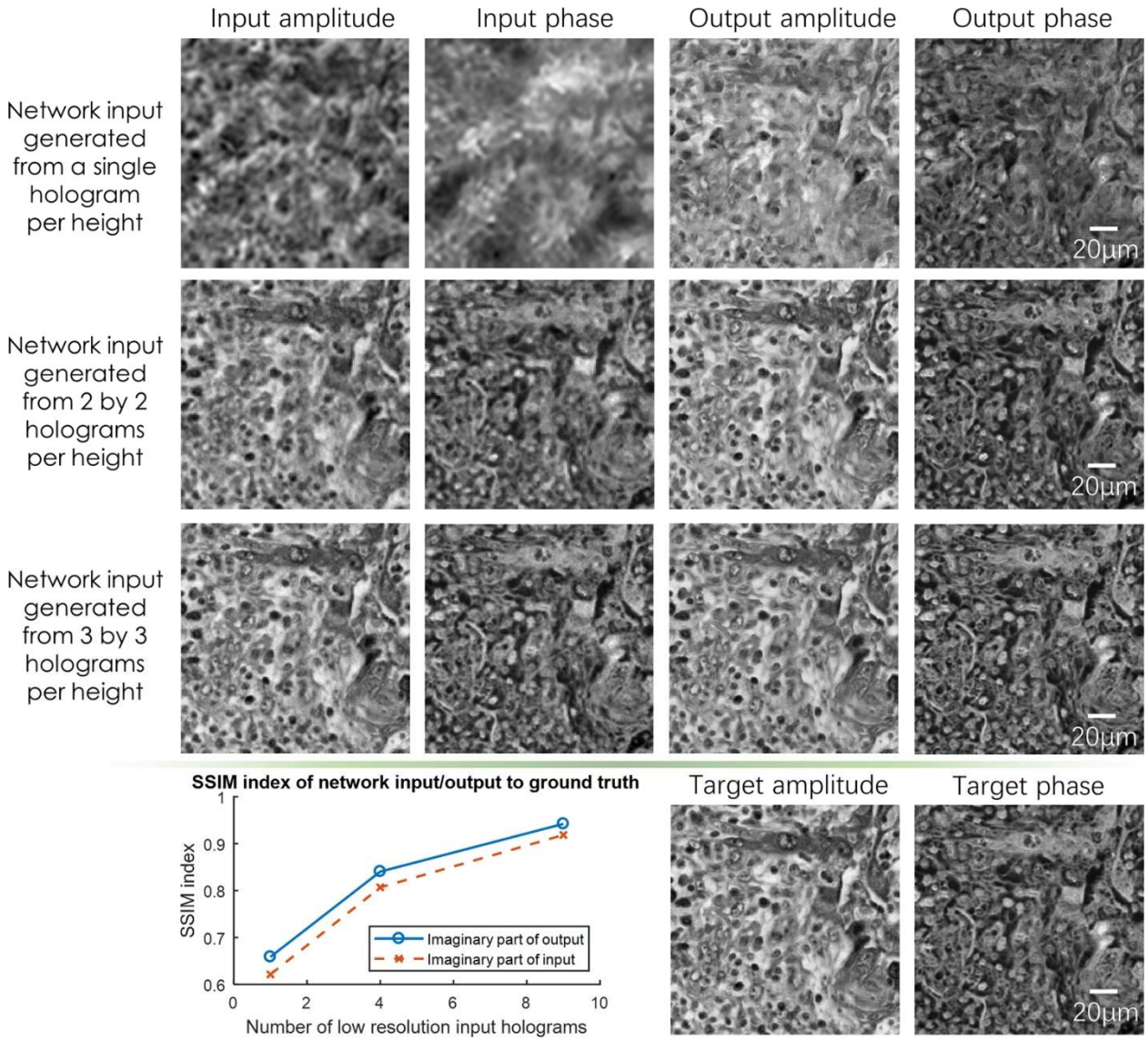

Figure 2.6 Visualized result for the pixel size-limited system. Comparison of the performances for the deep-learning-based pixel super-resolution methods using different input images. The sample is a Masson's trichrome stained lung tissue slide, imaged at an illumination wavelength of 550 nm. SSIM values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

I also report the SSIM values with respect to the reference label images in order to further evaluate the performance of the network output when applied to a pixel size-limited coherent imaging system. The average SSIM values for the entire image FOV (~20 mm^2) are listed in **Table**

2.3, where the input SSIM values were calculated between the bicubic interpolated lower resolution input images and the ground truth images. The results clearly demonstrate the improved structural similarity of the network output images.

Table 2.3 Average SSIM values for the lung and Pap smear samples for the deep neural network output (also see Figs. 5 and 6 for sample images in each category).

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

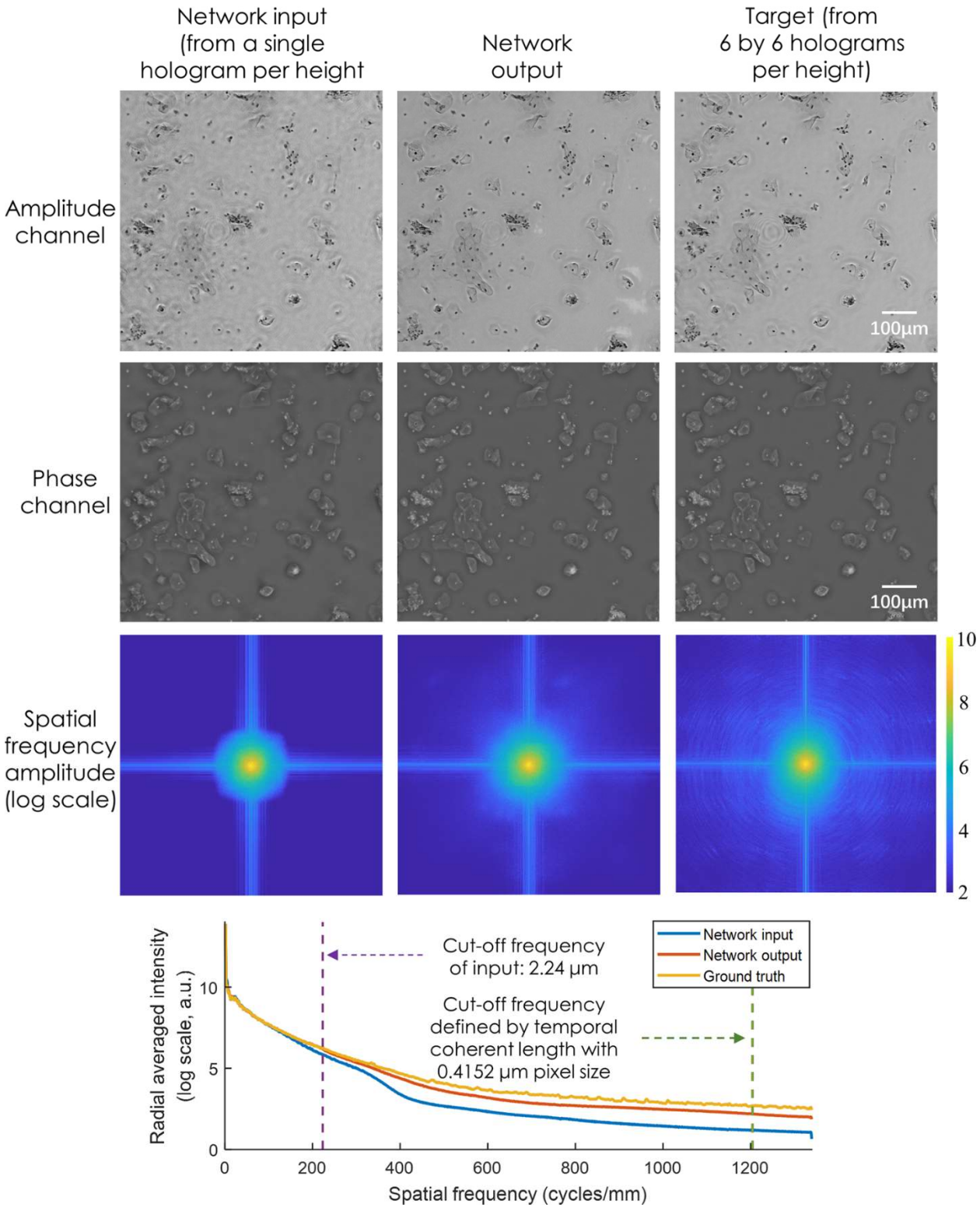

Figure 2.7 Spatial frequency analysis for the pixel size-limited system. Radially-averaged spatial frequency spectra of the network input, network output and target images, corresponding to a lensfree on-chip coherent imaging system.

In addition to SSIM, I also report the improved performance of the network output using spatial frequency analysis. **Figure 2.7** reports the 2-D spatial frequency spectra and the associated radially-averaged frequency intensity of the network input, network output and the ground truth images corresponding to the lensfree on-chip imaging system. Another indication that the super-resolution is successful is that the higher spatial frequency components in the output of the network are very close to the spatial frequencies of the ground truth image.

2.4 Super-resolution of a diffraction-limited coherent imaging system (System B).

To demonstrate that this super-resolution framework can also be applied to a diffraction-limited coherent imaging system, another network was trained using the same architecture (see the Methods section) with images taken from a Pap smear sample. As in the pixel super-resolution case reported earlier, two samples were obtained from two different patients, and the trained network was blindly tested on a third sample obtained from a third patient. The input images were obtained using a 4×/0.13 NA objective lens and the reference ground truth images were obtained by using a 10×/0.30 NA objective lens. **Figure 2.8** illustrates a visual comparison of the network input, output and label images, providing the same conclusions as in **Figure 2.5** and **2.6**. Similar to the pixel size-limited coherent imaging system, the performance of the network was analyzed using spatial frequency analysis, the results of which are reported in **Figure 2.9**. As in Fig.7, the higher spatial frequencies of the network output image approach the spatial frequencies observed in the ground truth images.

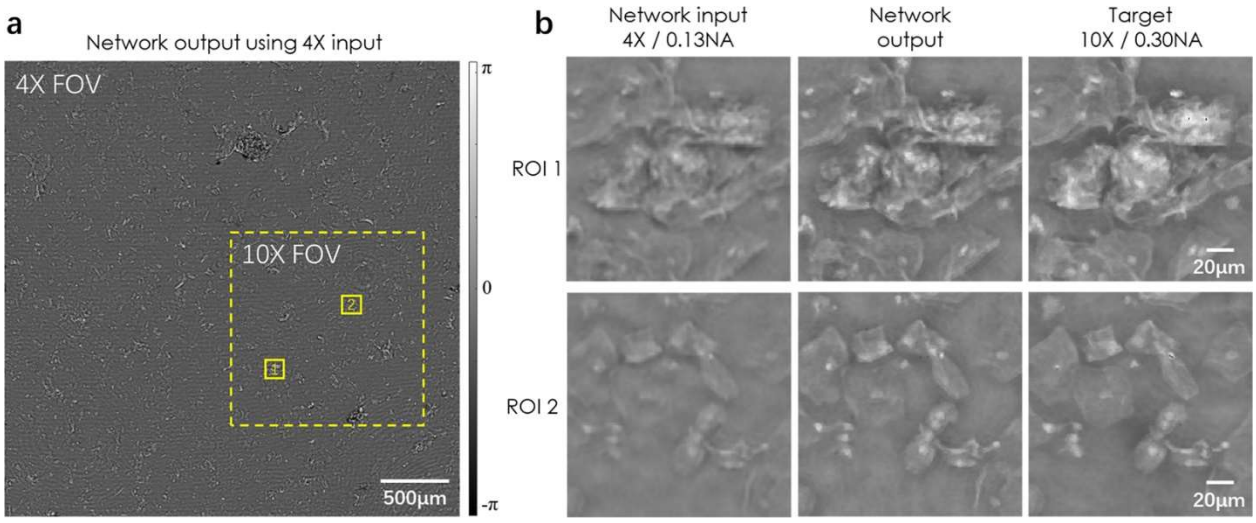

Figure 2.8 Visualized result for the diffraction-limited system. Deep learning-based super-resolution imaging of a Pap smear slide under 532 nm illumination using a lens-based holographic microscope. (a) Phase channel of the network output image. (b) Zoomed-in images of (a).

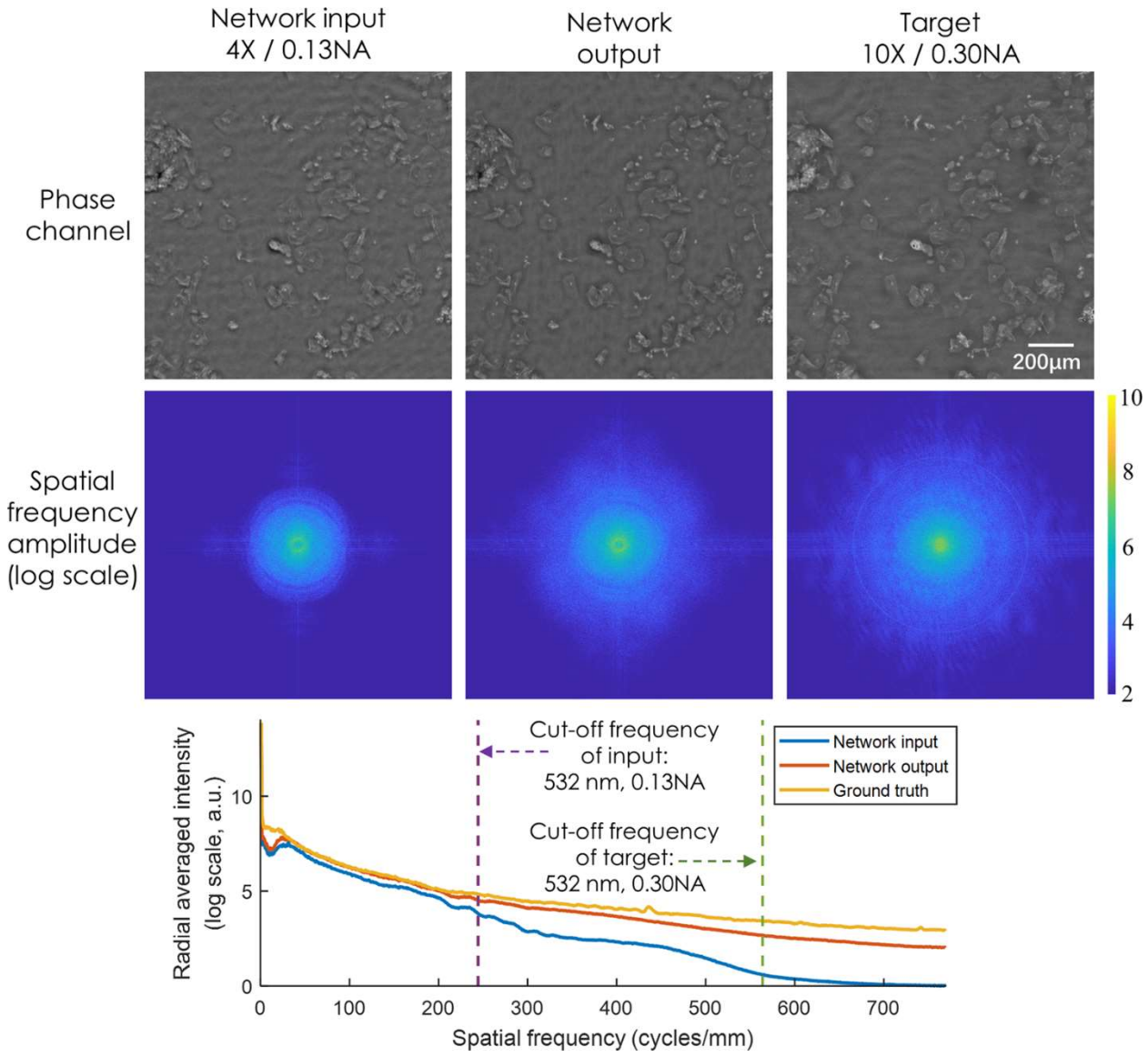

Figure 2.9 Spatial frequency analysis for the diffraction-limited system. Radially-averaged spatial frequency spectra of the network input, network output and target images, corresponding to a lens-based coherent imaging system.

The SSIM values for this system do not reveal as large of a trend as was observed for the lensfree on-chip microscopy system reported earlier. They only show a very small increase from a value of 0.876 for the input image to 0.879 for the network output. This is mainly due to increased coherence related artifacts and noise, compared to the lensfree on-chip imaging set-up. The lens-based design has several optical components and surfaces within the optical beam path, making it

susceptible to coherence induced background noise and related image artifacts, which can affect the SSIM calculations.

2.5 Conclusion

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

The proposed approach is not restricted to a specific coherent imaging modality and is broadly applicable to various coherent image formation techniques. The starting point for the super-resolution technique begins after the phase recovery step, which can be obtained by means of e.g., in-line, off-axis, phase-shifting holography, among other approaches. Since it has been proven to work for both pixel size-limited and diffraction-limited imaging systems, it will be applicable to other coherent systems with similar resolution limitations. One of the techniques that will highly benefit from the proposed framework is off-axis holography. The proposed technique might be used to bridge the space-bandwidth-product gap between off-axis and in-line coherent imaging systems, while retaining the single-shot and high sensitivity advantages of off-axis image acquisition systems.

Chapter 3 Super resolution of scanning electron microscope images

In this chapter I will demonstrate how a framework similar to the one in Chapter 2 can be used to enhance images coming from a scanning electron microscope. Electron microscopes are not limited in resolution by their wavelength, but rather the scanning pixel size and aberrations. Therefore, this chapter will only focus on pixel-limited super resolution. In this chapter I will show that this technique can be used to perform a 2× super resolution of electron microscope images and infer unresolved features that are visible only in high-resolution images coming from the same sample. Furthermore, I will also show that the method infers features which match the spatial frequency of those same high-resolution images.

Parts of this chapter have been published in:

- K. de Haan, Z.S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, "Resolution enhancement in scanning electron microscopy using deep learning," Scientific Reports DOI: 10.1038/s41598-019-48444-2 (2019).

3.1 Introduction

Scanning electron microscopy (SEM) is an important tool for characterization of materials at the nanoscale. By using electrons instead of photons for imaging samples, SEM can achieve sub-nanometer spatial resolution [52], revealing topological and compositional features invisible to traditional light microscopy. Therefore, SEM is frequently employed in a wide range of fields such as material science, biomedicine, chemistry, physics, nanofabrication, and forensics, among others [53–55]. For example, new applications such as nanocutting, where a silicon wafer can be cut at m/s speeds using a diamond blade have been demonstrated [56–59]. Applications such as

these require SEM characterization and therefore, as new tools to process or investigate the properties of silicon and other materials are developed, there will be an expanding need for improved electron microscopy tools.

However, when compared to light microscopy, the focused electron beam utilized by SEM is inherently more destructive to samples, especially soft and/or dielectric materials, resulting in electron charge build-up as well as deformation from absorption-based heating [60]. Consequently, these practical barriers prohibit many important samples such as biological specimens, polymers, and hydrogel-structures from being reliably characterized by SEM. There are, however, several approaches to mitigate the destructive effects of the electron beam. For example, it is common practice to coat the samples in e.g., gold, palladium, or iridium prior to imaging [61]. Additionally, shorter dwell times can be used during the electron beam scan to reduce the exposure to the sample. Though helpful, these approaches pose a performance trade-off: to reduce charging effects and sample deformation from heat one must alter the sample from its native state and/or incur increased noise in the acquired image [62].

Although computational approaches for super resolution in electron microscopy have been previously demonstrated [63,64], they require that a portion of the image be taken in high resolution or that the images have similar characteristics and contain sparse unique structures outside of a periodic topology. Other computational enhancements that have been applied to SEM images include denoising as well as deconvolution to reduce the spatial blur in the image caused by the finite beam size [65,66]. Alternative imaging techniques such as ptychography can also be used to increase the resolution in SEM, but these approaches require modification of the imaging set-up [67,68].

Here, I present a deep learning-based approach to improve the lateral resolution of SEM images using a neural network. By training a convolutional neural network (CNN) with a set of co-registered high- and low-resolution SEM images of the same set of samples, the network blindly super resolves individual SEM images, reducing sample charging and beam damage without losing image quality or adding extra sample preparation steps. In contrast to previous classical image enhancement methods, the approach can be implemented over a wide range of sample types, and only requires a single SEM image as input. Furthermore, by using deep learning to perform super-resolution, the need for prior knowledge of the image and approximated models that estimate the forward imaging operator is eliminated. Super-resolution using a single input is also advantageous as it simplifies image acquisition and therefore improves the speed and broadens the number of possible applications. Additionally, by using a co-registered and experimentally acquired training image dataset, the network can partially account for possible aberrations and noise in the imaging system. This data-driven approach has the added benefit of reducing the scanning time of the electron beam, and thus increasing the imaging throughput by enabling the use of a lower magnification scan over a larger field-of-view without sacrificing image quality.

Deep neural networks have emerged as an effective method for statistical processing of images and have been shown to improve image quality and achieve super resolution of camera images [7] and across several modalities of optical microscopy [8,41]. Once trained, the network can quickly process input SEM images in a feed-forward and non-iterative manner to blindly infer images with improved quality and resolution, thus making it an attractive and practical tool for rapid SEM image enhancement. Additionally, deep-learning based super-resolution has been proven to be more effective than other classical image enhancement techniques [7].

3.2 Imaging

The efficacy of this deep-learning based technique was demonstrated using a gold-on-carbon resolution test specimen [Ted Pella 617-a]. This test specimen has a random assortment of gold nanoparticles of varying sizes ranging from 5 nm to 150 nm immobilized on carbon, and is commonly employed to measure the resolution of SEM systems at different scales using the gaps between various gold nanoparticles. The network's effectiveness when applied to a hydrogel coated with a thin layer of gold is also demonstrated.

The image dataset employed to train the CNN was made up of unique high- and low-resolution pairs of the test specimen, each taken from the same region of interest where there is a distribution of nanoparticles. The low-resolution images were taken at a magnification of 10000× (14.2 nm pixel size), while the high resolution images were taken at 20000× magnification (7.1 nm pixel size). Higher magnification ground truth images (>20000×) were empirically found to lead to inaccurate inference results, and therefore the ratio of the label image magnification to the input image magnification was limited to 2. The training SEM image pairs were taken by first capturing the high resolution images, and then lowering the magnification and imaging the same field of view. In both cases the image resolution is limited by the number of pixels and therefore the lower magnification images can be modeled as aliased versions of the higher resolution images. A Nova 600 DualBeam-SEM (FEI Company) was used with a 10 kV accelerating voltage, 0.54 nA beam current, and a monopole magnetic immersion lens for high resolution imaging. All images were acquired with 30 μs pixel dwell time.

For the hydrogel imaging experiments, the low-resolution images were taken at a magnification of 2500× (56.8 nm pixel size), and the corresponding high resolution images were taken at 10000× magnification and binned to give an effective magnification of 5000× (28.4 nm pixel size). These images were acquired using a 10 μs pixel dwell time.

3.3 Methods

Co-registration

Once the high- and low-resolution image pairs were taken, they were co-registered before being inputted to the neural network for the training phase. These training images were first roughly matched to each other by cropping the center of each of the low-resolution images and using a Lanczos filter to up-sample the images. After this rough alignment, additional steps were taken to register the images with higher accuracy. First, image rotation and size misalignment were corrected by using the correlation between the two images to define an affine matrix which was then applied to the high resolution images. Next, local registration was performed using a pyramid elastic registration algorithm [38,69]. This algorithm breaks the images into iteratively smaller blocks, registering the local features within the blocks each time, achieving sub-pixel level agreement between the lower and higher resolution SEM images. The images were taken using automatic brightness and contrast adjustment. In order to account for possible fluctuations in the settings, both the high- and low-resolution images were normalized for use in the network training using the mean and standard deviation of the pixel values.

Network Training

For the gold nanoparticles, 40 pairs of accurately registered images (924×780 pixels) were split into 1920 non-overlapping patches (128×128 pixels) which were then used to train the network. The hydrogel image dataset was made up of 131 pairs of 492×418 pixel images, which were cropped into 4542 overlapping 128×128 pixel patches. 372 of these patches were automatically removed from the training dataset using an experimentally determined correlation threshold due to beam damage creating severe dissimilarities between the images. The sizes of the

training datasets were further increased by randomly rotating and flipping each image patch, and an identical network model was used for each dataset. The network model utilized in this chapter was a Generative Adversarial Network (GAN) which uses a generator network to create the enhanced images, and a discriminator network (D) that helps the generator network (G) to learn how to create realistic high-resolution images [11]. In addition to the standard discriminator loss, an L1 loss term was also added to ensure that the generated images are structurally close to the target, high-resolution images; the anisotropic total variation loss (TV) was also used to increase the sparsity of the output images and reduce noise. Based on this, the overall loss function for the generator network can be written as:

$$l_{generator} = L_1\{G(x), z\} + \alpha \times TV\{G(x)\} + \beta \times \left[1 - D\big(G(x)\big)\right]^2 \tag{3.1}$$

where x is the low resolution input image to the generator network and z is the matching high resolution ground truth image corresponding to the same field-of-view. α and β are tunable parameters to account for the relative importance of the different loss terms. The L_1 loss is the mean pixel difference between the generator's output and the ground truth image, defined as [70]:

$$L_1\{G(x), z\} = \frac{1}{M \times N} \sum_i \sum_j \left| z_{i,j} - G(x)_{i,j} \right| \tag{3.2}$$

where i and j are the pixel indices in an $M \times N$ pixel image. The anisotropic total variation loss is defined as [71]:

$$TV\{G(x)\} = \sum_i \sum_j \left(\left| G(x)_{i+1,j} - G(x)_{i,j} \right| + \left| G(x)_{i,j+1} - G(x)_{i,j} \right| \right) \tag{3.3}$$

The discriminator loss, on the other hand, penalizes the discriminator when it is unable to discriminate between the generated and the ground truth images, and is defined as [11]:

$$l_{discriminator} = D\big(G(x)\big)^2 + \big(1 - D(z)\big)^2 \tag{3.4}$$

The discriminator loss, L1 loss, and the total variation loss make up 84%, 14%, and 2% of the total loss for the generator, respectively. The generator uses an adapted U-net structure [4], while the discriminator uses a modified Visual Geometry Group (VGG) type network structure [2]. Details of these network architectures are shown in **Figure 3.1**.

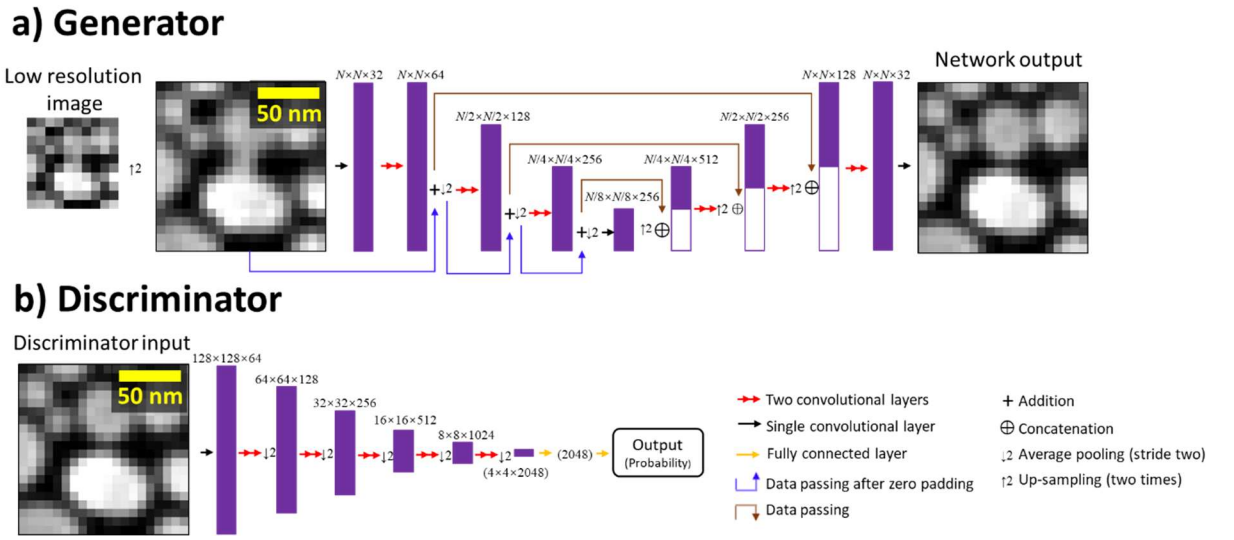

Figure 3.1 Diagram of the network structure. Every convolutional block is made up of two convolutional layers, each followed by a leaky rectified linear unit (ReLU) activation function. The second convolutional layer in each block changes the number of channels. a) The structure of the generator portion of the network. b) The structure of the discriminator portion of the network.

The network was implemented in Python (version 3.6.2) using the TensorFlow library (version 1.8.0). For both samples, the generator was trained for 48,000 iterations with the discriminator updating every fourth iteration to avoid overfitting. This took the network one hour and twenty minutes to train using a single Nvidia GTX 1080 Ti graphics processing unit (GPU)

and an Intel Core i9-7900 processor. The same computer is able to infer 3.66 images per second, for an image size of 780×780 pixels. This inference time is 16 times faster than the low-resolution SEM imaging of the corresponding sample field-of-view; stated differently, real-time visualization of the super-resolved images, immediately after a low-resolution image is taken or while a new scan is ongoing, is feasible.

3.4 Results

This super resolution technique allows us computationally to enhance the resolution of lower magnification SEM images such that the network's output accurately matches the resolution given by the higher resolution SEM images of the same samples. A demonstration of this can be seen in **Figure 3.2**, which reports several blindly tested examples of nanoparticles that are not clearly resolved in the input images, but become distinct after the application of the neural network. These fields of view are distinct from those used to train the network, but taken from different areas of the same sample. Pixel-intensity cross-sections are also reported to illustrate the resolution enhancement more clearly. From these examples we can see that the network is able to reveal spatial details that are not clear in the input (lower magnification) SEM images, matching at its output the corresponding higher magnification SEM images of the same fields-of-view. This is particularly evident in the gaps between the gold nanoparticles shown in **Figure 3.2**.

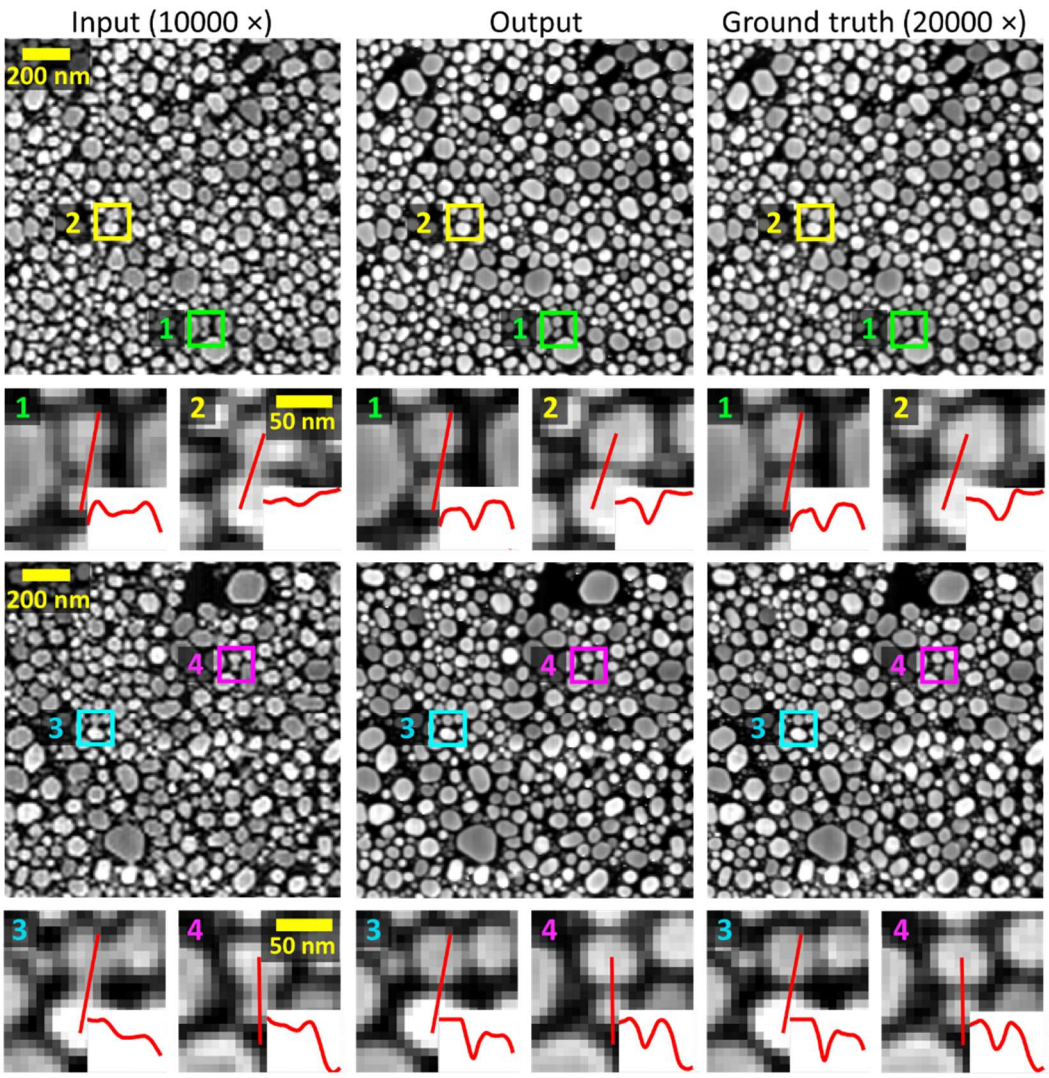

Figure 3.2 Examples of the up-sampled network input images compared to the output and ground truth SEM images. Cross sections of various spatial features with noticeable resolution enhancement are shown.

In fact, **Figure 3.3** provides a statistical analysis of these gaps to quantify the enhancement provided by the trained network; for this analysis, 300 gaps between arbitrary adjacent nanoparticles were randomly selected using the high-resolution SEM images. They were then analyzed to determine whether the neighboring particles are resolvable, as well as to quantify the gap-size in the input image, output image, and target image. The gap width was defined as the distance between the points at which the intensity drops below 80% of highest intensity value of

the adjacent particles, and a gap was determined to exist if the lowest intensity point between the particles fell below 60% of the peak value. In the input SEM image (lower magnification), 13.9% of these gaps were not detectible, i.e., could not be resolved (see Fig. 3). However, after super resolving the input SEM images using the trained network, the percentage of undetected gaps dropped to 3.7%. Additionally, the *average absolute difference* between the measured gap sizes in the low- and high-resolution SEM images decreases from 3.8 nm to 2.1 nm after passing through the network.

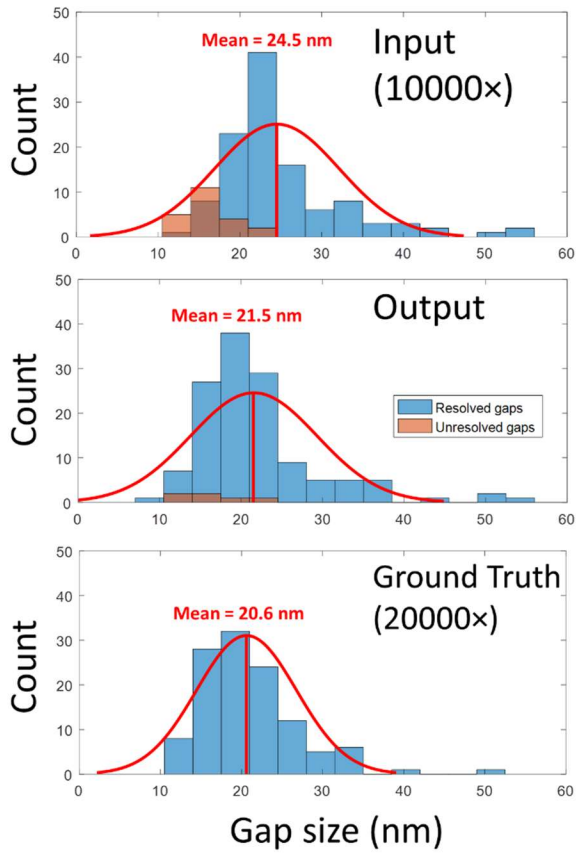

Figure 3.3 Histograms of the gap sizes inferred from the network input and the output images compared to the ground truth image. Total count changes among the histograms due to some of the gaps only being visible in specific images. In the input SEM images, 13.9% of the gaps were not detectible; the percentage of undetected gaps dropped to 3.7% for the output images. A Gaussian distribution, fitted to the gap histograms, with the corresponding mean gap size is also shown for each plot. The number of unresolved gaps in both the input and output images is also shown using a different color; unresolved gaps were not used for mean gap estimation. Pixel size per image is 7.1 nm; the input image is upsampled by a factor of 2.

Another way to illustrate the resolution improvement is reported in the spatial frequency analysis shown in **Figure 3.4**. This figure compares the magnitudes of the spatial frequencies for the low- and high-resolution SEM images as well as those of the network output images. From this comparative analysis we can see that the network enhances the high frequency details of the input SEM image such that the spatial frequency distribution of the network output image is consistent

with the high-resolution SEM image – including the spatial frequencies that are aliased in the input image due to the large pixel size.

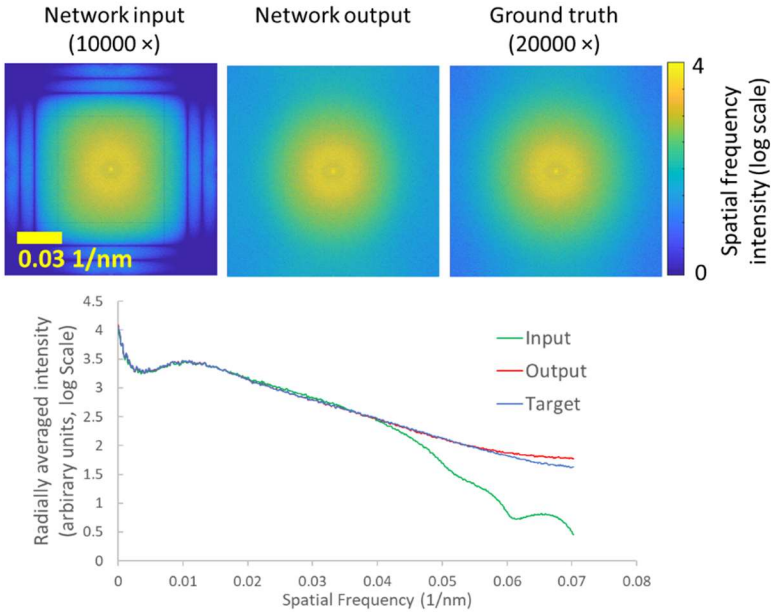

Figure 3.4 Top: spatial frequency distributions of the average of five up-sampled input, output, and ground truth images are compared. Bottom: radially-averaged plot of the above distributions. Analysis was performed on the uncropped versions of the SEM images shown in **Figure 3.2** as well as three additional images.

To demonstrate the performance of the network when applied to a sample that is prone to beam damage, a hydrogel sample was used. In this case, the network was able to improve the image quality by sharpening and denoising it (**Figure 3.5**). However, the electron beam-induced damage of the sample during imaging makes accurate co-registration impossible and therefore reduces the success of the network training process and its inference. An example of this can be seen in **Figure 3.5**. In cases like this, the performance of the super-resolution network will decrease as the high- and low-resolution images used to train the network become more dissimilar due to movement of the specimen. However, once a network has been trained, the inference process

requires only a single input image. Therefore, this approach can help to mitigate some of the sample deformation caused by the radiation as the low-resolution image can be captured using reduced beam intensity.

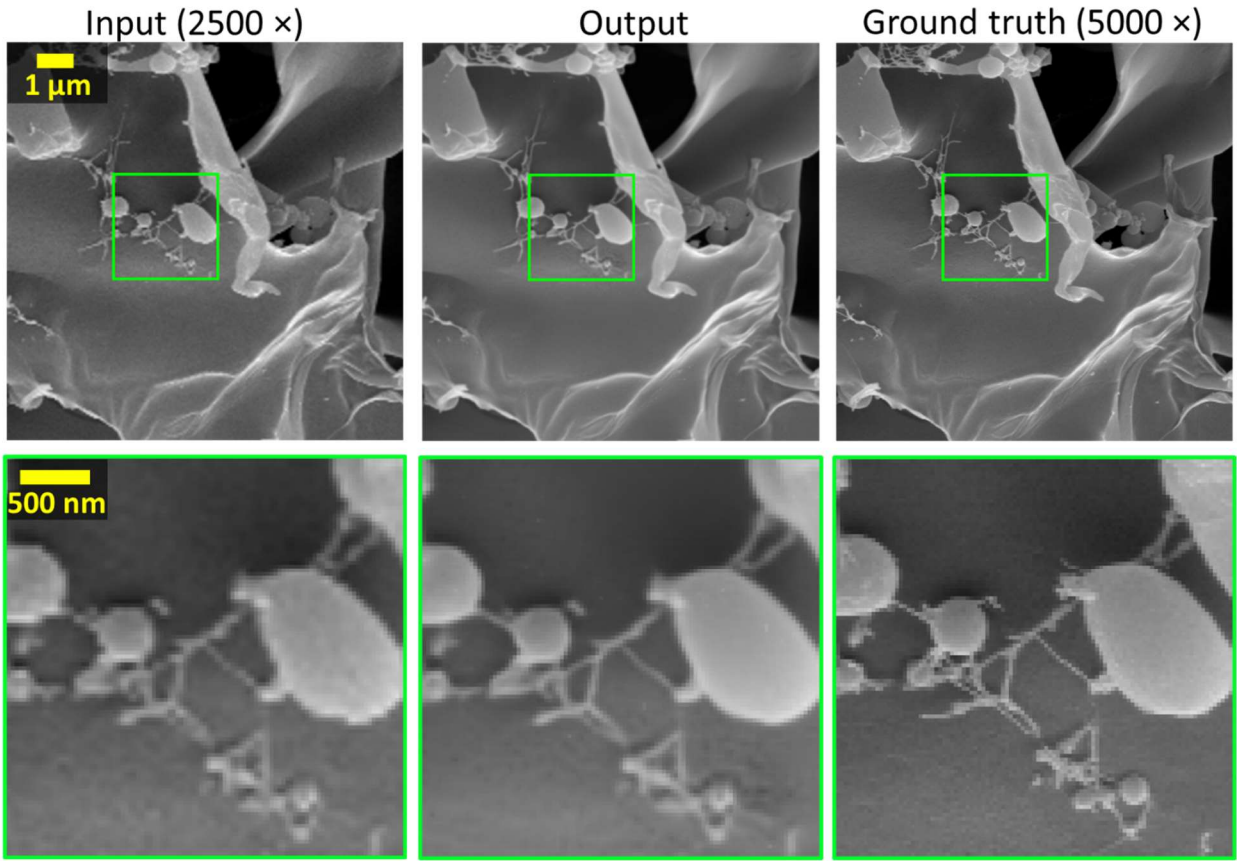

Figure 3.5 Examples of the up-sampled hydrogel network input images compared to the network output and the ground truth SEM images. Top: full field of view. Bottom: zoomed in region of interest. As the hydrogel sample is partially damaged during the imaging process, the ground truth images have a slightly different structure than the input images. This sample damage makes subpixel image co-registration impossible, which reduces the quality of the network training and limits the success of its inference, compared to earlier presented results.

3.5 Conclusion

Taken together, deep learning-based super resolution is shown to be a powerful and practical tool to computationally improve the resolution in SEM. The 2-fold increase in resolution

demonstrated here allows for a *four-fold* reduction of the number of electrons which must interact with the sample to acquire an SEM image, in turn enabling a four-fold increase in the speed of image acquisition. While the demonstrated method is less effective when trained using samples that are prone to significant beam damage or other movement during the acquisition of the training dataset, it could benefit the characterization of samples that present limited charging or beam-induced damage by reducing the electron exposure without sacrificing the image quality. This would allow for higher resolution imaging of a variety of biological materials and nanofabricated samples that previously could not be characterized adequately by SEM. While I demonstrated the effectiveness of this network in the ideal case, where the training and testing datasets are made up of the same type of sample, deep learning-based super-resolution techniques have been shown to generalize to other types of samples or magnification factors, where larger datasets were used or there were strong similarities between different types of samples [8]; this can be further improved through transfer learning [41], which can be performed in almost real time by acquiring a few calibration images corresponding to a new type of sample of interest.

Chapter 4 Automated detection of sickle cell disease using enhanced smartphone microscope images

Beyond simply improving the resolution of microscopy images, this deep learning-based framework can be used to correct for a wide variety of aberrations ranging from color correction to minor changes in focus. In this chapter, I will present an application this technology, and use it to enhance brightfield images of thin blood smears captured by a mobile phone microscope to match the quality of an industry standard benchtop scanning microscope. By improving the image quality and standardizing the images, it allows a second neural network to be applied to the enhanced images, performing semantic segmentation and detecting sickle cells and normal red blood cells found within the images. In turn, the slide can have an automated sickle cell disease (SCD) diagnosis performed using this segmentation. I will show that this combination of neural networks can be used to accurately diagnose sickle cell disease in 96 patients (32 having SCD), with an accuracy of ~98% and an area under the curve of 0.998. In addition to standardizing images for the classification network, the enhancement network improves images captured by a smartphone microscope to a level which could be used by a diagnostician if required.

Parts of this chapter have been published in:

- K. de Haan, H.C. Koydemir, Y. Rivenson, D. Tseng, E. Van Dyne, L.S. Bakic, D. Karinca, K. Liang, M. Ilango, E. Gumustekin, and A. Ozcan, "Automated screening of sickle cells using a smartphone-based microscope and deep learning," npj Digital Medicine DOI: 10.1038/s41746- 020-0282-y (2020).

This was reused from ref. [72] under the Creative Commons Attribution 4.0 International License.

4.1 Introduction

Sickle cell disease is the most common hematologic inherited disorder worldwide and a public health priority [73]. The majority of the world's SCD burden is in Sub-Saharan Africa, affecting millions of people at all ages. It is estimated that 200,000 to 300,000 children are born with SCD every year in Africa alone [74,75]. The prevalence of the disease varies across countries, being approximately 20% in Cameroon, Ghana and Nigeria and even rising up to 45% in some parts of Uganda [75].

SCD is an inherited disorder caused by a point mutation in hemoglobin formation, which causes the polymerization of hemoglobin and distortion of red blood cells in the deoxygenated state. As a result of this, the normally biconcave disc-shaped red blood cells become crescent or sickle shaped in people living with SCD. These red blood cells are markedly less deformable, have one-tenth the life span of a healthy cell, and can form occlusions in blood vessels. Children with SCD also suffer from spleen auto-infarction and the burden of disease becomes significant. Loosing splenic function, these children are at high risk for infections at an extremely young age, which significantly increases mortality rates [76]. Due to the lack of diagnosis and treatment, over 50% of these of children with SCD in middle and low-income countries will die [77].

Various methods have been developed for screening and diagnosis of SCD, including e.g., laboratory-based methods such as high performance liquid chromatography (HPLC) [78], isoelectric focusing [79], and hemoglobin extraction [80]. In addition to these relatively costly laboratory-based methods, there have been SCD diagnostic tests developed for point-of-care (POC) use [81–86]. These POC tests are mainly based on human reading, and human errors along with the storage requirements of these tests (involving e.g., controlled temperature and moisture to preserve chemical activity/function) partially limit their effectiveness to screen SCD, especially in resource limited settings [86].

An alternative method used for screening of SCD involves microscopic inspection of blood smear samples by trained personnel. In fact, each year hundreds of thousands of blood smear slides are prepared in sub-Saharan Africa to make diagnosis of blood cell infections and disorders [87]. Peripheral blood smears, exhibiting variations in e.g., the size, color, shape of the red blood cells can provide diagnostic information on blood disorders including SCD [88]. In addition to diagnosis, inspection of blood smears is also frequently used for evaluation of treatment and routine monitoring of patients [89]. Preparation of these blood smear slides is rather straight-forward (i.e., can be performed by minimally-trained personnel), rapid and inexpensive. However, this method requires a trained expert to operate a laboratory microscope and perform manual analysis once the blood smear is prepared; the availability of such trained medical personnel for microscopic inspection of blood smears is limited in resource scarce settings, where the majority of people with SCD live [90]. In an effort to provide a solution to this bottleneck, deep learning-based methods have been previously used to classify [91] and segment [92] different types of red blood cells from digital images that were acquired using laboratory-grade benchtop microscopes equipped with oil-immersion objective lenses. However, these earlier works focused upon cell level detection, rather than slide level classification and therefore did not demonstrate patient level diagnosis or screening of SCD.

As an alternative to benchtop microscopes, smartphone-based microscopy provides a cost-effective and POC-friendly platform for microscopic inspection of samples, making it especially suitable for use in resource limited settings [93–95]. Smartphone microscopy has been demonstrated for a wide range of applications, including e.g., the imaging of blood cells [96,97], detection of viruses and DNA [98,99], quantification of immunoassays [100–103] and microplates [104] among many others [105–108]. Recently, machine learning approaches have

also been applied to smartphone microscopy images for automated classification of parasites in soil and water [109,110].

Here I present a smartphone-based microscope and machine learning algorithms that together form a cost-effective, portable and rapid sickle cell screening framework, facilitating early diagnosis of SCD even in resource-limited settings. This mobile microscope (**Figure 4.1**) utilizes an opto-mechanical attachment, coupled to the rear camera of a smartphone, transforming it into a portable microscope using external parts that cost ~$60 in total. This compact microscope design has sub-micron spatial resolution [38] and weighs only 350 g including the smartphone itself.

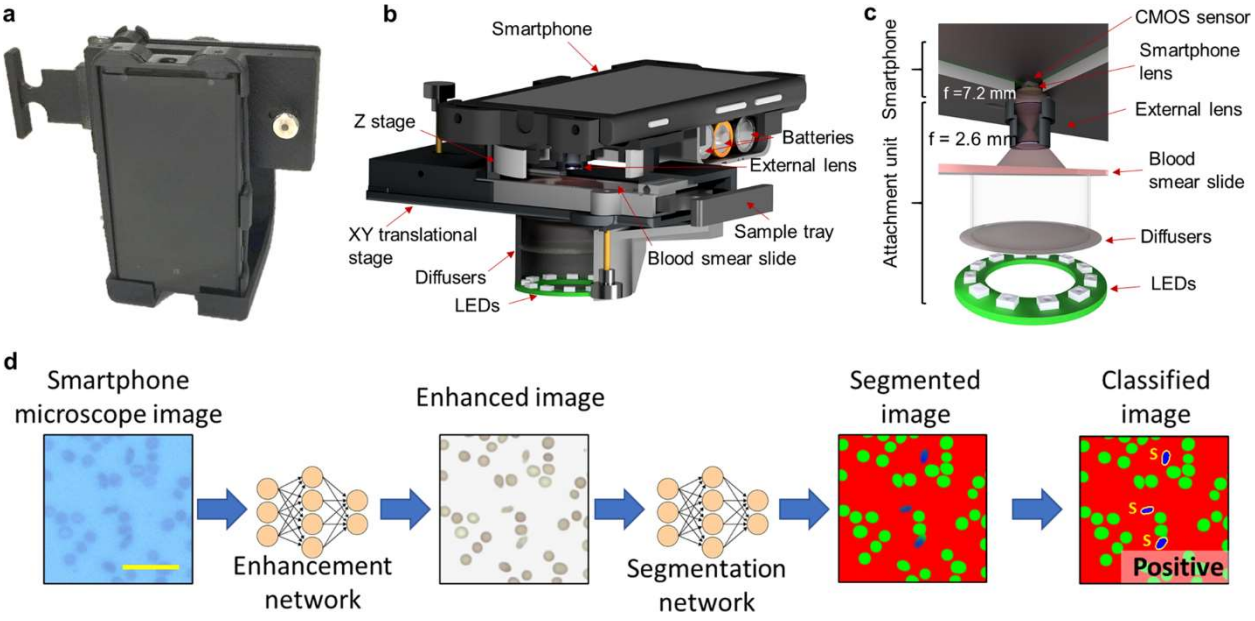

Figure 4.1 Field portable smartphone based brightfield microscope and its principle of detection. a) A photograph of the smartphone-based brightfield microscope. Schematic illustration of b) the design of the portable microscope in detail and c) the light path. d) Deep learning workflow for sickle cell analysis. Scale bar indicates 20 μm.

Using this cost-effective mobile microscope, slide-level automated diagnosis of SCD was performed by rapidly classifying thousands of red blood cells within a large field-of-view using a deep learning-based framework that takes <7 sec to process a blood smear slide per patient. The

approach was blindly tested using 96 blood smears (32 of which came from individual patients with SCD) and achieved ~98% accuracy together with an area-under the-curve (AUC) of 0.998. Therefore, this platform provides a robust solution for cost-effective and rapid screening of SCD, making it especially promising for POC use in resource-limited settings.

4.2 Results

The process of screening peripheral blood smears using this framework is illustrated in **Figure 4.1**(d). Following the imaging of the patient slide with the smartphone-based microscope (**Figure 4.1**(a, b, and c)), these smartphone images were rapidly enhanced using a deep neural network as shown on the left part of **Figure 4.1**(d). This network was trained to transform the lower resolution, spatially and spectrally aberrated images of the smartphone microscope into enhanced images that are virtually equivalent to the images of the same samples captured using a higher numerical aperture (NA) laboratory-grade benchtop microscope. During the training phase (Chapter 4.4, Methods section), which is a one-time effort, pairs of mobile microscopy images (input to the network) were cross-registered to the corresponding images of the same training samples, captured using a ×20/0.75 NA objective-lens with a ×2 magnification adapter (i.e., ×40 overall magnification), which were used as ground truth image labels.

This intermediate image transformation is essential for not only the automated blood smear analysis using a subsequent classification neural network (**Figure 4.1**(d)), but is also important for the improvement of smartphone microscopy images to a level that can be used by expert diagnosticians for manual inspection of the blood smears. Due to aberrations and lower resolution, the raw smartphone microscope images of red blood cells might be relatively difficult to interpret by human observers, who are traditionally trained on high-end benchtop microscopes. While this framework automatically performs cell classification and slide-level SCD diagnosis, any manual

follow-up by a trained expert requires digital images that can be accurately interpreted. This is an important need satisfied by the image enhancement neural network that is specifically trained on blood smear samples to enhance the smartphone microscope images.

Another major advantage of this approach is that the enhancement network *standardizes* the microscopic images of blood smears, making it easier for the second classification network perform its task and more accurately classify the sickle cells. Therefore, the enhancement network improves both the *quality* and the *consistency* of the subsequent sickle cell classification network by performing an image standardization at its output. It also helps us to account for variations between images over time. In fact, the sample collection was performed over the course of 3 years, and the blood smear images were captured with different smartphones (using the same opto-mechanical attachment, same phone manufacturer and model number), resulting in some variations between the acquired mobile-phone images over time. In addition to using multiple smartphones, variations can be caused by the 3D printed plastic attachments partially warping over time, the smartphone heating up during its operation, and the illumination intensity fluctuating over time. These changes can cause fluctuations in the alignment of the lenses to occur, which in turn lowers the quality of the captured images. The image enhancement network was trained with blood smear samples imaged over this time period and in different operational conditions in order to account for these variations in the raw image quality of smartphone microscopy, successfully standardizing the resolution and the spatial as well as spectral features of the output images of the network.

Using the structural similarity index (SSIM) [111], I quantified the effectiveness of this image enhancement network on unique fields-of-view from the same slides that the network was trained with. The neural network improved the SSIM of the smartphone microscopy images from 0.601±0.017 (input) to 0.965±0.012 (output), where the mean and standard deviation were

calculated for 8 smartphone microscope images, compared with the ground truth image labels acquired with a high NA benchtop microscope (see the Methods Section). This shows that after the neural network-based transformation, the intensity component of the smartphone images become highly similar to the benchtop microscope images. Examples of these image transformations can be seen in **Figure 4.2**, where direct comparisons between the network output and the ground truth benchtop microscope (0.75 NA) images are shown.

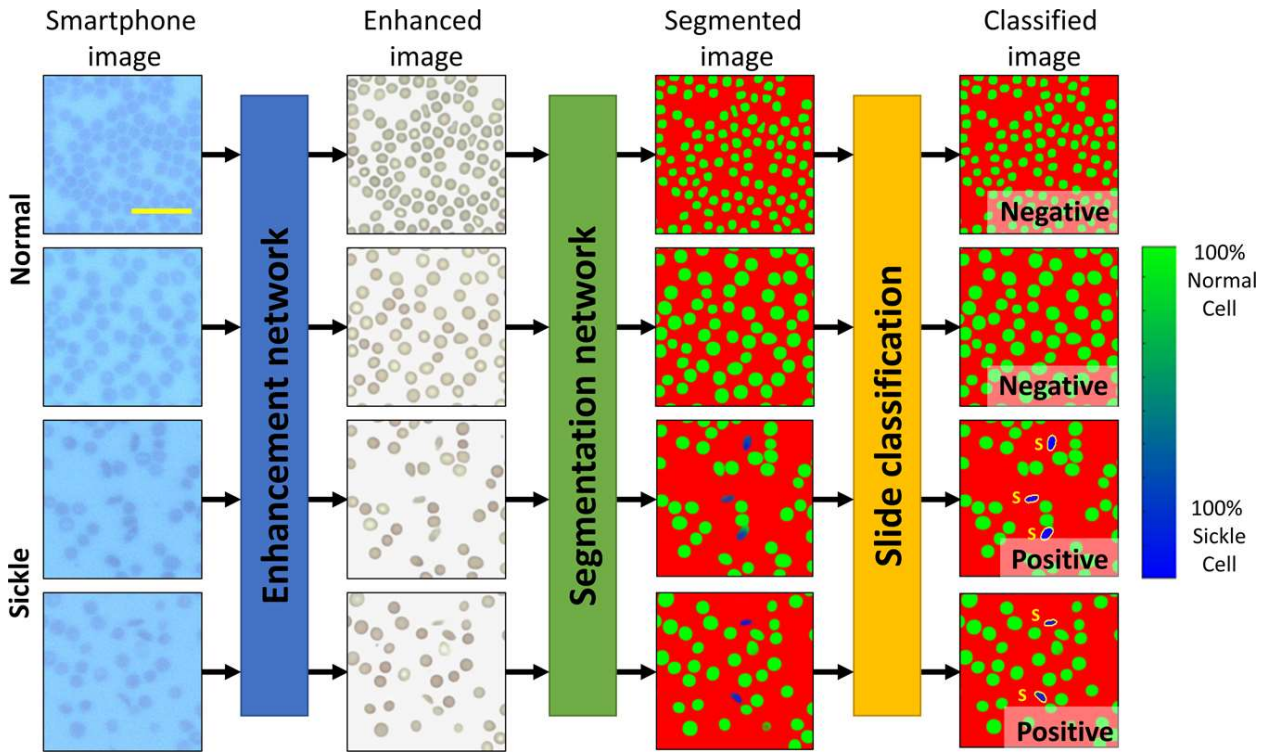

Figure 4.2 Example image patches. Demonstration of various test image patches that passed through the various steps of the automated sickle cell analysis framework. The smartphone microscopy images are first passed through an image enhancement and standardization network. Following this step, the images are segmented using a second, separate neural network. This segmentation network is in turn used to determine the number of normal and sickle cells within each image; 5 fields-of-view together covering ~1.25 mm^2 is automatically screened, having on average 9630 red blood cells to make a diagnosis for each patient blood smear. Scale bar indicates 20 μm.

Following the image enhancement and standardization network output, a second classification network was then used to segment the enhanced images into three classes of objects: normal red blood cells, sickle cells, and background. Using the output of this network, each patient blood smear slide imaged by smartphone microscopy was automatically analyzed and screened for SCD using 5 different fields-of-view, each covering 0.51 mm × 0.51 mm, i.e., a total of ~1.25 mm^2 area of each blood smear was processed by the classification network, screening on average 9630 red blood cells per patient sample. Following this segmentation, the number of sickle cells and normal cells contained within each image were automatically counted. The patient slides were classified as being SCD positive if the average number of sickle cells within 5 images covering a 1.25 mm^2 field-of-view was above 0.5% of the total red blood cell count for that sample. This threshold was chosen based on the performance of the classification neural network in the validation dataset (separate from the blind testing data) to mostly account for sickle shaped healthy cells found in normal blood smears.

After the training phase, the accuracy of the framework was confirmed by blindly testing 96 blood smear slides that were never seen by the networks before. Covering 96 individual patients, 32 of these blood smears are from SCD patients and 64 of them are from healthy individuals. These slides were imaged using the smartphone microscope between 2016 and 2019, and were anonymously obtained from existing specimen at the UCLA Medical Center; the clinical diagnosis of each patient sample was used as the ground truth label for each slide. In this blinded test, this framework achieved ~98% accuracy across these 96 blood smears, where there was one false positive slide and one false negative slide. As for the misclassifications, one healthy blood smear was found to have a significantly higher average number of sickle cells (0.64%) than the remaining healthy blood smears; the one false negative sample was only slightly below the 0.5% threshold,

while exhibiting a higher percentage of sickle cells than any of the remaining normal blood smears. Several examples of patches from these test images, as they digitally pass through the networks, are also shown in **Figure 4.2**.

I also report the receiver operating characteristic (ROC) curve of the framework in **Figure 4.3**, which demonstrates how the SCD diagnosis accuracy can change depending on the threshold used to label the blood smear slide, achieving an AUC of 0.998.

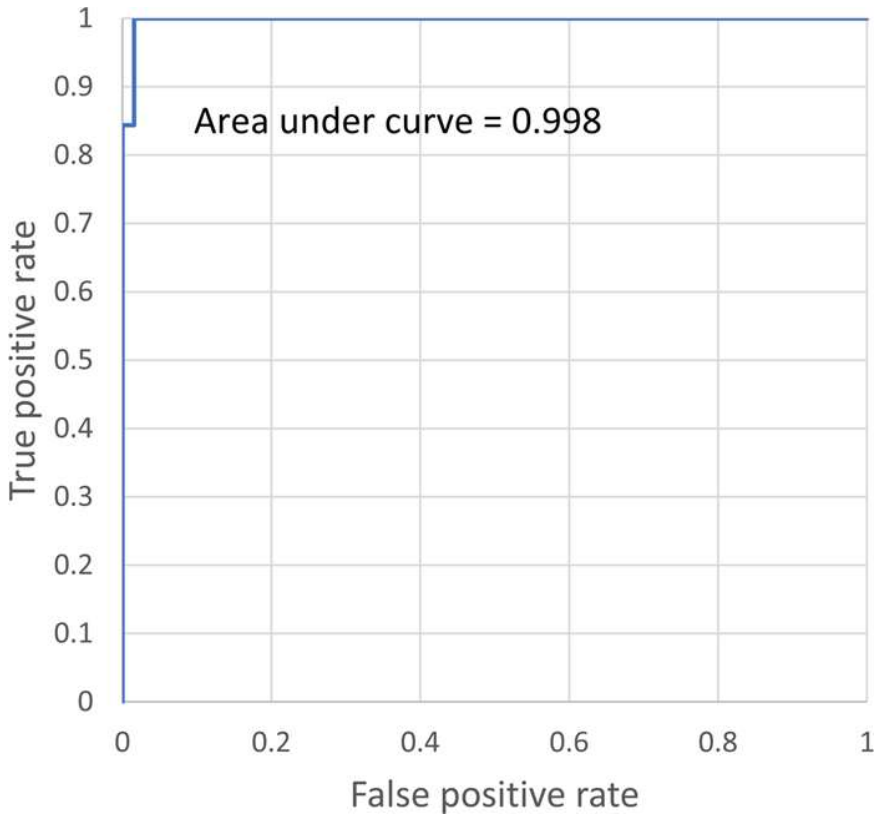

Figure 4.3 ROC curve. Demonstration of the false positive rate versus the true positive rate for the sickle cell detection framework.

4.3 Discussion

Through blind testing spanning 96 unique patient samples I have shown that the presented framework can consistently achieve high accuracy even using a limited training dataset. Similar

to human diagnosticians examining blood smears under a microscope, screening through a large number of cells (on average 9630 red blood cells per patient sample) helped us achieve a high accuracy for automated diagnosis of SCD. In general, by using more training data containing a larger number of labeled sickle cell images, one can further improve the SCD detection accuracy and make it more efficient, requiring the capture of fewer smartphone images covering an even smaller field-of-view. On the other hand, acquiring a large training dataset with accurate labels at the single cell level is a challenge in itself; in fact, single cell level ground truth labels from blood smear samples clinically do not exist, and are not being recorded. Although one could perform single cell level molecular analysis on blood samples of patients, creating a sufficiently large scale image library backed up with such single cell level ground truth labels would be very costly and time consuming. As an alternative, one can use multiple diagnosticians to establish a clinical ground truth at the single cell level by statistically merging the decisions made by a panel of diagnosticians. A similar multi-diagnostician based decision approach has previously been used to detect malaria infected cells in blood smears, rather than relying on a single expert [112,113]. This approach mitigates the fact that single cell level labeling which is performed by a human can be highly subjective and inconsistent even for highly-trained experts [112]. Therefore, an accurate patient diagnosis can be difficult to perform with only a limited number of cells screened per slide, particularly with diseases such as SCD, where normal/healthy blood can also contain cells showing sickle cell like microscopic features. For example, children with normal blood were shown to have on average 0.28% sickle cells [114]. Due to this variation, a large number of cells must be screened per patient slide to better evaluate the rate of occurrence and make an accurate diagnosis.

Given these aforementioned challenges in obtaining large scale ground truth labeled image data from blood smears, the image enhancement and standardization network is particularly vital

for SCD screening, helping the subsequent red blood cell segmentation network to better generalize sickle cell features and be efficient with limited training data by standardizing the input images that are fed into the classification network.

In this work, the blood smear slides were classified by measuring the percentage of sickle cells over a field-of-view of ~1.25 mm^2, covering on average 9630 red blood cells. **Figure 4.4**(a) reports how the diagnostic accuracy and AUC would change as a function of the number of cells that are screened per patient, further demonstrating the importance of inspecting a large sample field-of-view, and therefore a large number of cells for accurate SCD diagnosis (see the Methods section). Without inspecting an average of a few thousand red blood cells per patient slide, the accuracy of the automated SCD screening platform can drop significantly. **Figure 4.4**(b) also reports how the ROC curves are impacted as a function of the number of cells being screened per patient slide, confirming the same conclusion that both the sensitivity and the specificity of the test steadily drop as the number of inspected cells decreases.

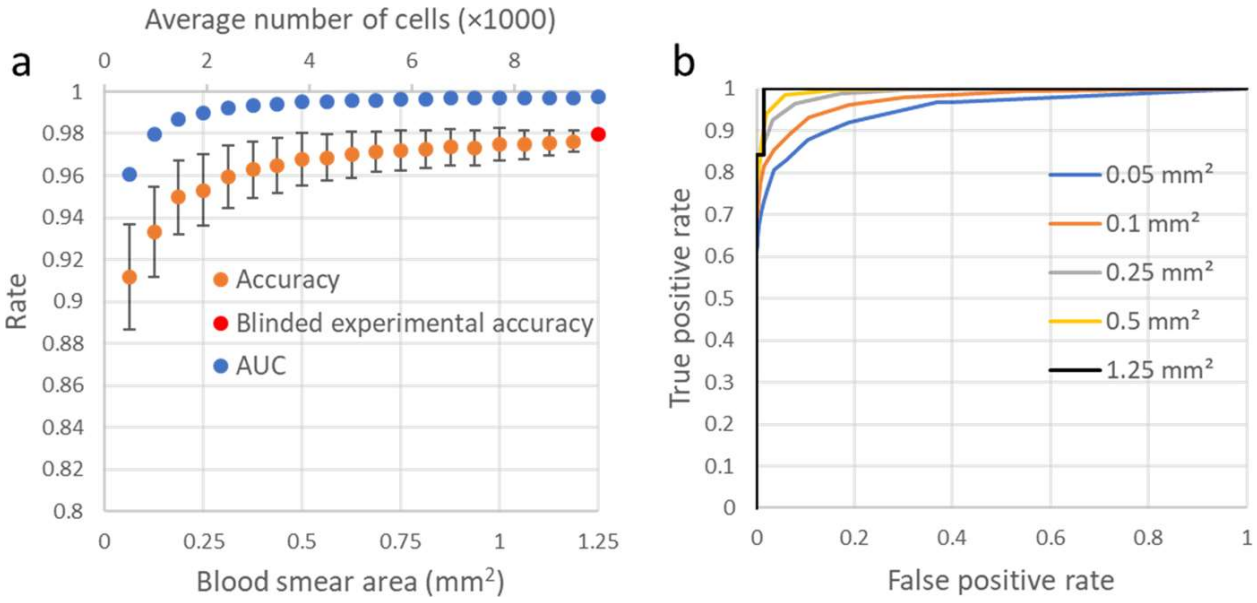

Figure 4.4 Accuracy as a function of the number of cells counted. a) Plot of how the accuracy and AUC change as a function the number of cells (and the blood smear area) inspected by the method. b) ROC curves for various simulated blood smear areas. These plots (except the 1.25 mm^2 one, which is the experimental result) are based on the average of 1000 Monte Carlo simulations performed by removing the red blood cells from the imaging fields-of-view at random to change the number of cells inspected by the method. As the cells are relatively monodisperse, this random removal of the cells simulates a reduction in the inspected blood smear area per patient. Error bars represent the standard deviation (s. d.) across the 1000 Monte Carlo simulations.

These results and analysis demonstrate that the presented method, enabled by smartphone microscopy and deep learning, is robust to perform SCD diagnosis by automated processing of blood smears. However, there are also some drawbacks to using a blood smear to diagnose sickle cell disease which the automated system is unable to overcome. One of these is that using blood smear slides to diagnose sickle cell disease or sickle cell anemia is not effective for infants and can cause false negatives. Therefore, the method presented in this chapter is not to be used in newborn screening programs, and has been designed for use with older children and adults. While newborn screening programs are rather important, there are many areas where they have not been implemented due to e.g., cost and logistical issues. Furthermore, even in areas where newborn

screening programs have been implemented, there are still adults and children who have not been screened due to a lack of supplies, power outages, or lack of complete coverage, which makes a mobile phone-based cost-effective SCD screening technology ideal to serve such patients. Blood smears also cannot be used to distinguish between sickle cell genotypes, and therefore patients may need follow up testing. However, for critical clinical management of e.g., an initial episode of stroke, chest syndrome, or hand-foot syndrome, it may not be clinically necessary to know the sickle cell genotype immediately.

Therefore, this technology would ideally be targeted towards adults and older children who have not undergone newborn screening. I believe that it would be particularly useful in settings where the existing point of care technologies are not ideal due to cost, need for reagents etc. as the presented SCD screening method is rapid, cost-effective and the required sample preparation is minimal. It would be particularly useful in locations where operators are already trained to use thin blood smears to perform diagnoses of other diseases. The preparation of blood smears is being routinely performed in resource limited settings for the diagnosis of other diseases such as malaria (thick/thin smears) and complete blood counts (when specialized machines are not available), resulting in hundreds of thousands of blood smears prepared per year just in sub-Saharan Africa [115]. This method is also quite fast to compute an answer: each one of the 5 images passes through the neural networks in 1.37 seconds using a single Nvidia 1080 Ti GPU. This results in a total analysis time of 6.85 sec per patient, which is significantly faster than any manual inspection by experts. While in the current implementation the analysis is performed on a computer (which can be a local or remote server), a smartphone application could be also used to perform the processing on the phone itself with an increase in the slide processing time.

4.4 Methods

Study design

The objective of the research was to perform automated diagnosis of SCD using blood smear slides imaged with a smartphone-based microscope and analyzed by deep learning. The test dataset was made up of 96 unique patient samples involving 32 SCD thin blood smears and 64 normal thin blood smears. The blood smears were obtained from the UCLA Ronald Reagan Hospital, and no sample selection criterion was applied. Both the normal and SCD blood smears came from patients with a mix of gender and wide range of ages (<5 years to >60 years old).

Ethics approval and consent to participate

De-identified blood smears from existing human samples were obtained from the UCLA Ronald Reagan Hospital. No patient information, or any link to it, was disclosed to the research team. General consent for the samples to be used for research purposes was obtained. Due to these factors, no specific IRB from UCLA was required.

Design of the smartphone-based brightfield microscope

The Nokia Lumia 1020 smartphone attached to a custom-designed 3D-printed unit was used to capture images of the blood smear slides. An RGB light emitting diode (LED) ring (product no. 1643, Adafruit) was utilized to illuminate the sample in a transmission geometry and a microcontroller (product no. 1501, Adafruit) was used to adjust the illumination intensity. To ensure uniform illumination of the sample, a diffuser (product no. SG 3201, American Optic Supply, Golden, CO, USA) was placed in between the LEDs and the specimen. An external lens with a focal length of 2.6 mm (LS-40166 – M12xP0.5 Camera Lens) was used for magnification and was coupled to the rear camera of the smartphone. This design has a field of view of ~ 1 mm^2 per image. However, each one of the mobile phone images is cropped to the center ~0.5x0.5 mm^2

in order to avoid increased aberrations toward the edges of the field of view; per patient slide, 5 independent images were captured, covering a total of ~1.25 mm^2. The microscope is also equipped with a custom-designed manual translational stage to laterally move the sample. This stage, which was also 3D-printed, also contained a microscope slide holder. In total, the smartphone-based microscope weighs 350 g, including the phone itself, and the total cost of the device parts is ~ $60 (excluding the smartphone).

Imaging of thin blood smears

Thin blood smear slides were used for image analysis. The ground truth microscope images were obtained using a scanning benchtop microscope (model: Aperio Scanscope AT) at the Digital Imaging Laboratory of the UCLA Pathology Department. The standard smartphone camera application was used to capture the corresponding input images using the smartphone-based microscope, using auto focus, ISO 100, and auto exposure.

Areas of the samples captured using the smartphone microscope were co-registered to the corresponding fields-of-view captured using the benchtop microscope (please refer to "Image co-registration" in Methods section for details). Three board-certified medical doctors labelled the sickle cells within the images captured using the benchtop microscope using a custom-designed graphical user interface (GUI). As the images are co-registered, these labels were used to mark the locations of the sickle cells within the smartphone images, forming the training image dataset. The images were captured on the feathered edges of the blood smear slides, where the cells are dispersed as a monolayer.

Images from blood smears containing cells which have been scraped and damaged were excluded from the dataset, as the cut cells can appear similar to sickle cells. One normal blood

smear was accordingly excluded as we were unable to capture a sufficient number of usable fields-of-view due to the poor quality of the blood smear, with many scratches on its surface. Blood smears from four patients who were tested positive for SCD and were taking medicine for treatment were also excluded from the study since their smears did not contain sickle cells when viewed by a board-certified medical expert.

Image co-registration

The co-registration between the smartphone microscope images and those taken by the clinical benchtop microscope (NA=0.75) was done using a series of steps. For the first step, these images are scaled to match one another by bicubically down-sampling the benchtop microscope image to match the size of those taken by the smartphone. Following this, they are roughly matched using an algorithm which creates a correlation matrix between each smartphone image and the stitched whole slide image captured using the benchtop lab-grade microscope. The area with the highest correlation is the field of view which matches the smartphone microscope image and is cropped from the whole slide image. An affine transformation was then calculated using MATLAB's (Release R2018a, The MathWorks, Inc.) multimodal registration framework which extracts feature vectors and matches them to further correct the size, shift, shear, and account for any rotational differences [116]. Finally, the images were matched to each other using an elastic pyramidal registration algorithm to match the local features [38]. This step accounts for the spherical aberrations, which are extensive due to the nature of the inexpensive optics coupled to the smartphone camera. This algorithm co-registers the images at a subpixel level by progressively breaking the image up into smaller and smaller blocks and uses cross-correlation to align them.

Image enhancement neural network

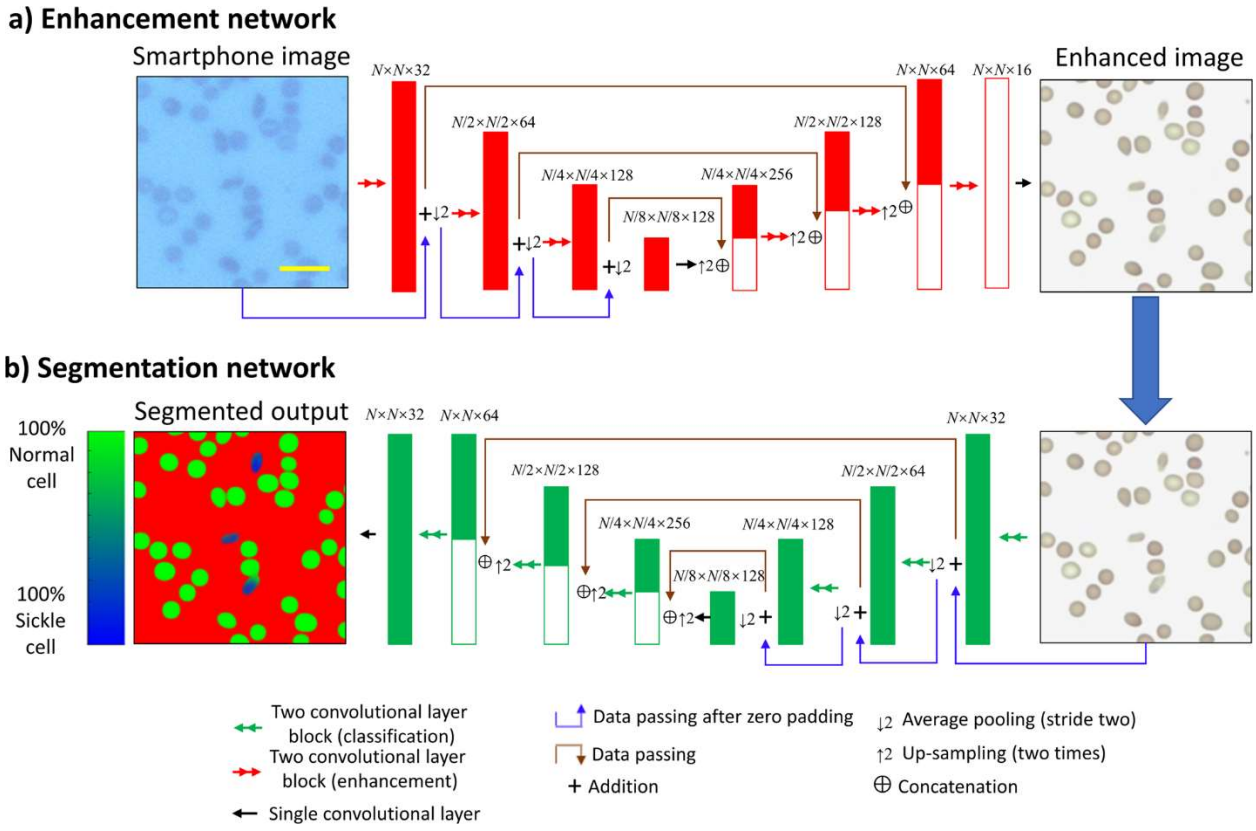

Figure 4.5 Deep neural network architecture. Diagram detailing the network architecture for both a) the image enhancement network and b) the semantic segmentation network. The numbers above the layers represent the size of the tensor dimensions at that point in the network, in the format: length × width × number of channels. N was chosen to be 128 during training and can be set to any arbitrary size during testing. Scale bar indicates 20 μm.

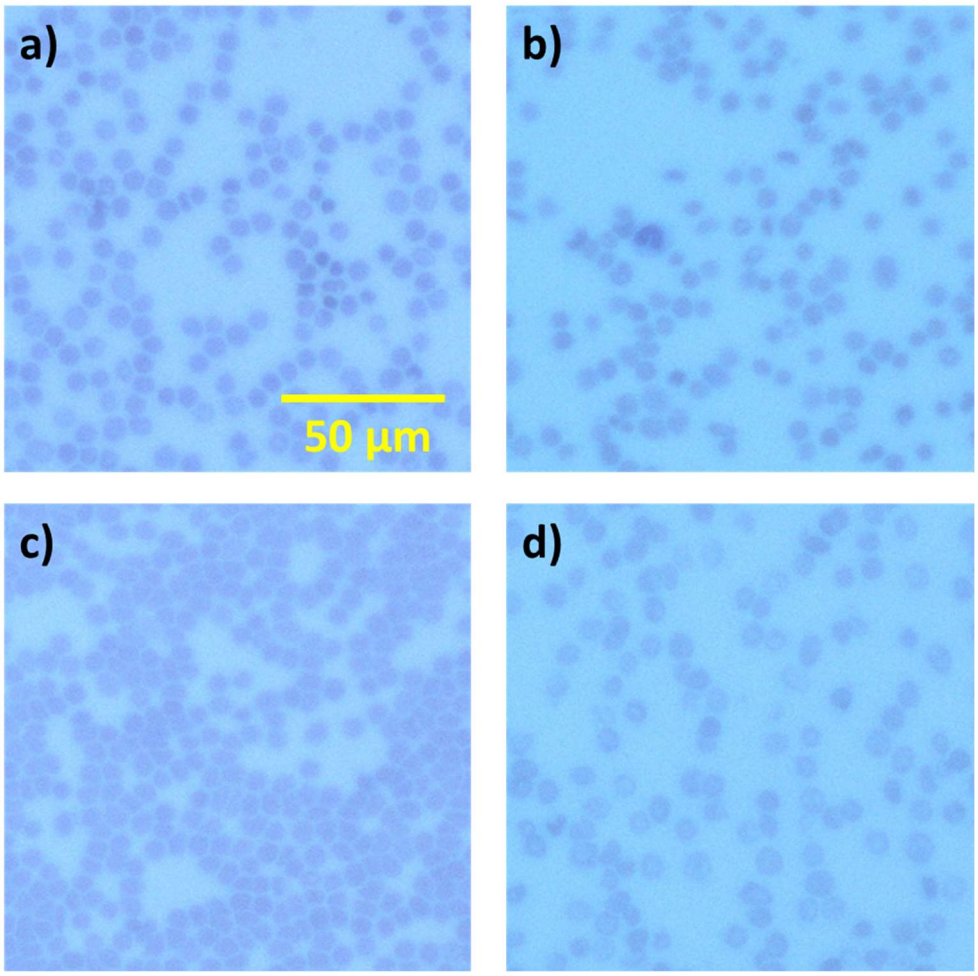

Figure 4.6 Demonstration of the aberrations and various differences among smartphone microscope images. Aberrations include blur due to the device's inherent challenge to accurately focus on the sample, as well as blur in certain areas of the image due to the usage of cost-effective optical components, tilt and defocus aberrations due to optical misalignments, spectral aberrations between the fields of view as well as between the cells within each field of view, and pixel saturation in the centers of some of the cells (e.g., the cells in panel (d)).

Due to the variations among the images taken by the smartphone microscope, a neural network is used to standardize images and improve their quality in terms of spatial and spectral features. These variations stem from e.g., changing exposure time, aberrations (including defocus), chromatic aberrations due to source intensity instability, mechanical shifts, etc. Some examples of the image variations that these aberrations create can be seen in **Figure 4.6**. The quality of the

images taken by a smartphone microscope can be improved and transformed so that they closely resemble those taken with a state-of-the-art benchtop microscope by using a convolutional neural network [38]. The image normalization and enhancement network uses the U-net architecture as shown in **Figure 4.5**(a) [4]. The U-net is made up of three "down-blocks" followed by three "up-blocks". Each one of these blocks is made up of three convolutional layers, which use a zero padded 3×3 convolution kernel and a stride of one to maintain the size of the matrices. After each of the convolutional layers, the leaky ReLU activation function is applied, which can be described as:

$$\text{Leaky ReLU (x)} = \begin{cases} x & \text{where } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \tag{4.1}$$

where x is the tensor that the activation function is being applied to.

In the case of the down-block, the second of these layers increases the number of channels by a factor of two, while the second convolutional layer in the up-block reduces the number of channels by a factor of one quarter. The down-blocks are used to reduce the size of the images using an average pooling layer with a kernel size and a stride of two, so that the network can extract and use features at different scales. The up-blocks return the images to the same size by bilinearly up-sampling the images by a factor of two. Between each of the blocks of the same size, skip connections are added to allow information to pass by the lower blocks of the U-net. These skip connections concatenate the up-sampled images with the data from the down-blocks, doubling the number of channels. As the up-blocks reduce the number of channels by a factor of four and the skip connections double the number of channels, the total number of channels in each subsequent up-block is halved. Between the bottom blocks, a convolution layer is also added to allow processing of those large-scale features. The first convolutional layer of the network initially

increases the number of channels to 32, while the last one reduces the number back to the 3 channels of the RGB color space to match the benchtop microscope images (ground truth). The network was trained using the adaptive movement estimation (Adam) optimizer with a learning rate of 1×10^{-4}.

The image enhancement network is trained using a combination of two loss functions, described by the equation:

$$L_{Network} = L_1\{z, G(x)\} + \lambda * TV\big(G(x)\big) \tag{4.2}$$

where an L_1 (mean absolute error) loss function is used to train the network to perform an accurate transformation, while the total variation (TV) loss is used as a regularization term. λ is a constant set to 0.03; this constant makes the total variation ~5% of the overall loss. *G(x)* represents image generated using the input image *x*. The L_1 loss can be described by the following equation [70]:

$$L_1\{z, G(x)\} = \frac{1}{N_{\text{channels}} \times M \times N} \sum_{n=1}^{N_{\text{channels}}} \sum_{i,j=1}^{M,N} \left|G(x)_{i,j,n} - z_{i,j,n}\right| \tag{4.3}$$

Where $N_{channels}$ is the number of channels, *n* is the channel number, *M* and *N* are the width and height of the image in pixels, and *i*, and *j* are the pixel indices. The total variation loss is described by the following equation [71]:

$$TV(G(x)) = \frac{1}{N_{\text{channels}} \times M \times N} \sum_{n=1}^{N_{\text{channels}}} \sum_{i,j=1}^{M,N} \Big(\left|G(x)_{i+1,j,n} - G(x)_{i,j,n}\right| + \left|G(x)_{i,j+1,n} - G(x)_{i,j,n}\right|\Big) \tag{4.4}$$

The network was trained for 604,000 iterations (118.5 epochs) using a batch size of 16. The data were augmented through random flips and rotations of the training images by multiples of 90 degrees.

For this image enhancement network training, there is *no need* for manual labeling of cells by a trained medical professional, and therefore this dataset can be made diverse very easily. Because of this, it can also be expanded upon quickly as all that is required is additional images of the slides to be captured by both microscopy modalities and co-registered with respect to each other. Therefore, the network was able to more easily cover the entire sample space to ensure accurate image normalization and enhancement. The training image dataset consisted of 520 image pairs coming from 10 unique blood smears. Each of these images have 1603×1603 pixels, and are randomly cropped into 128×128 pixel patches to train the network. Several examples of direct comparisons between the network's output and the corresponding field of view captured by the benchtop microscope can be seen in **Figure 4.7**(a).

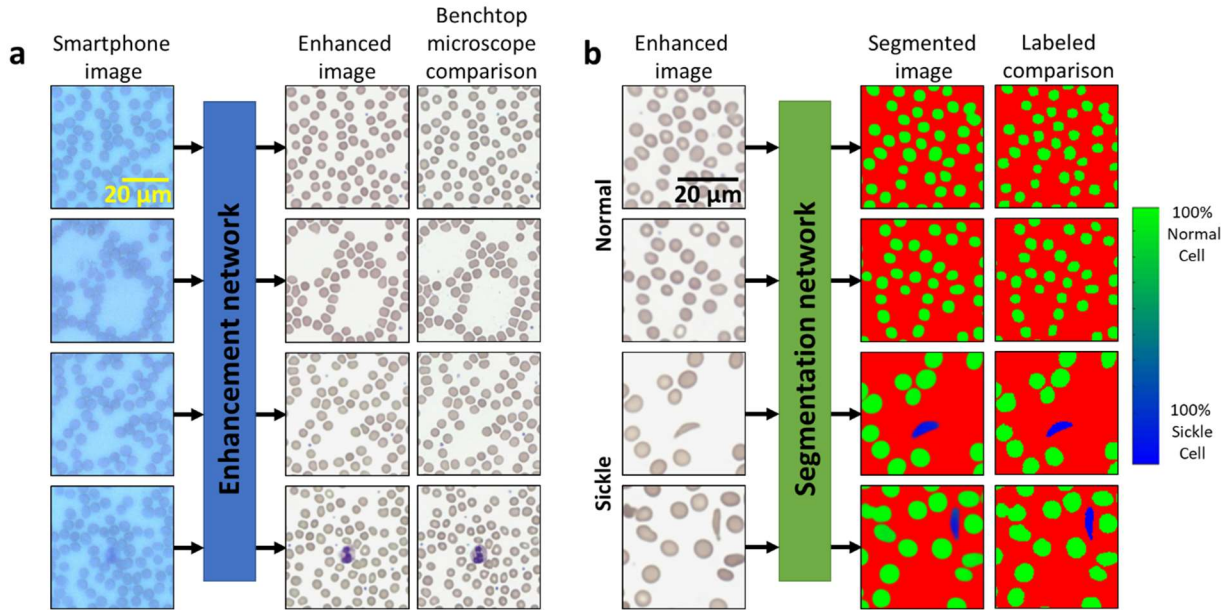

Figure 4.7 Direct comparison between the network output and the ground truth for the two networks. a) Demonstration of the image enhancement network, where the smartphone image is the network input, the enhanced image is the network output, and the benchtop microscope (NA=0.75) is the ground truth image label. b) Demonstration of the segmentation network, where the enhanced image is the network input, a segmented probability map is the network output, and the probability mask based on the expert's label is the ground truth.

Mask creation for training the cell segmentation network

Once the cells were labeled by board-certified medical experts and the images were co-registered, the cell labels were used to create a mask which constitutes the ground truth of the segmentation network; this mask creation process is a one-time training effort and used to train the cell segmentation neural network used in this work. These training masks were generated by thresholding the benchtop microscope images according to color and intensity to determine the locations of all the healthy and the sickle cells. The exact thresholds were chosen manually for each slide due to minor color variations between the blood smears; once again, this is only for the training phase. As the centers of some red blood cells were the same color as the background, holes in the mask were filled using MATLAB's imfill command, a morphological operator. Following this, the mask was eroded by four pixels in order to eliminate sharp edges and eliminate pixels misclassified due to noise. Any cell labeled by the medical expert as a sickle cell was set as a sickle cell while any unlabeled red blood cell was set as a normal cell for training purposes. White blood cells, platelets and the background were all labeled as a third background class. As the medical experts might have randomly missed some sickle cells within each field of view, a 128×128 region around each labeled sickle cell was cut out of the slide for training, reducing the unlabeled area contained within the training dataset. The remaining sections of the labeled slides were removed from the training dataset. At the end of this whole process, which is a one-time training effort, three classes are defined for the subsequent semantic segmentation training of the neural network: (1) sickle, (2) normal red blood cell, and (3) background.

Semantic segmentation

A second deep neural network is used to perform semantic segmentation of the blood cells imaged by the smartphone microscope. This network has the same architecture as the first image

enhancement network (U-net). However, as this network performs segmentation, it uses the SoftMax cross entropy loss function to differentiate between the three classes (sickle cell, normal red blood cell, and background). In order to reduce the number of false positives as much as possible, the normal cell class is given twice the weight of the background and the sickle cells in the loss function. The overall loss function for the segmentation network, $L_{Segmentation}$, is described in equation (4.5):

$$L_{Segmentation} = -\frac{1}{M \times N} \sum_{i,j=1}^{M,N} a_{i,j,1} \log\left(p_{i,j,c=1}\right) + 2a_{i,j,2} \log\left(p_{i,j,c=2}\right) + a_{i,j,3} \log\left(p_{i,j,c=3}\right) \quad (4.5)$$

where M and N are the number of pixels in an image, and i, and j are the pixel indices as above. $a_{i,j,c}$ is the ground-truth binary label for each pixel (i.e., 1 if the pixel belongs to that class, 0 otherwise), and c denotes the class number (c={1,2,3}), where the first class represents the background, the second class is for healthy cells, and the third class is for sickle cells. The probability $p_{i,j,c}$ that a class c is assigned to pixel i,j is calculated using the softmax function:

$$p_{i,j,c} = \frac{\exp(y_{i,j,c})}{\sum_{k=1}^{3} \exp(y_{i,j,k})} \quad (4.6)$$

where y is the output of the neural network.

A visual representation of the network architecture can be seen in **Figure 4.5**(b). Several examples of direct comparisons between the network's output at the single cell level and the corresponding field of view imaged by the clinical benchtop microscope can be seen in **Figure 4.7**(a).

This network was trained for 80,000 iterations (274 epochs) using a batch size of 20. The training dataset for this network was made up of 2660 sickle cell image patches (each 128×128

pixels) from a single blood smear slide, each one containing a unique labeled sickle cell. An additional 3177 image patches (each 128×128 pixels) coming from 15 unique slides containing solely normal cells were also used. Separate from the blind testing image dataset which involved 96 unique patients, 250 labeled 128×128-pixel sickle cell image patches and two 1500×1500-pixel images from healthy image slides were used as validation dataset for the network training phase. The classification algorithm was validated using these images alongside 5 unique fields-of-view from 10 additional blood smear slides of healthy patients. The network was trained using the Adam optimizer with a learning rate of 1×10^{-5}.

Classification of blood smear slides

Once the images have been segmented by the second neural network, the number of total cells and sickle cells must be extracted. The algorithm first uses a threshold to determine which pixels are marked as cells. Areas where the *sum* of the sickle cell and normal cell probabilities is above 0.8 are considered to be part of a red blood cell, while areas below this threshold are considered as background regions. Connected areas which contain more than 100 pixels above the 0.8 threshold are then counted to determine the total number of cells. Sickle cells are counted using a similar methodology: connected areas where there are over 100 pixels above a sickle cell probability threshold of 0.15 were counted as sickle cells. This threshold is set to be low since significantly more number of healthy red blood cells is used to train the network. A slide is classified as being positive for sickle cell disease when the percentage of sickle cells among all the inspected cells (sickle and normal red blood cells) over a total field-of-view of ~1.25 mm^2 is above 0.5%. The 0.5% threshold was chosen using the validation image dataset, i.e., it was based on the network's performance in classifying the 10 healthy validation slides to account for false positives and the occurrence of sickle shaped cells in normal blood smears. Several examples of direct comparisons

between the network's output and the ground truth labels for blindly tested regions of the labeled slides are shown in **Figure 4.7**(b).

Structural similarity calculations

The SSIM calculations were performed using only the brightness (Y) component of the YCbCr color space as the intensity contrast component are expected to remain similar, while the chroma components (Cb, Cr) to depend on other factors, including variability in the slide's staining. The color difference between the smartphone microscope images and the benchtop microscope images is also significant. The smartphone microscope images appear with a blue background, and should not directly be compared against the benchtop microscope images in the RGB color space. Therefore, using a color space where the brightness component can be extracted separately is necessary.

The calculations were performed upon 8 unique fields-of-view from the same slides which were used to train the enhancement network. SSIM is calculated using the equation:

$$SSIM(x,z) = \frac{(2\mu_x\mu_z + c_1)(2\sigma_{x,z} + c_2)}{(\mu_x{}^2 + \mu_z{}^2 + c_1)(\sigma_x{}^2 + \sigma_z{}^2 + c_2)} \tag{4.7}$$

where x and z represent the two images being compared, as above. μ_x and μ_z represent the average values of x and z respectively, and σ_x and σ_z are the variance of x and z, and σ_z is the covariance of x and z. c_1 and c_2 are dummy variables, which stabilize the division from a small denominator.

Monte Carlo simulation details

The Monte Carlo simulations reported in **Figure 4.4** demonstrate how the accuracy of the presented technique changes as a function of the number of cells analyzed by the neural networks; these simulations were implemented by beginning with the full cell count from the 5 fields-of-

view tested for each patient slide. This total cell count was reduced by randomly eliminating some of the cells to evaluate the impact of the number of cells analyzed on the accuracy. As the cells are relatively monodisperse, this random removal of red blood cells was used as an approximation of a reduction of the inspected blood smear area per patient. The results of 1000 simulations were averaged since the accuracy can fluctuate significantly, particularly at low numbers of cells. The total number of cells within the 5 fields-of-view that were used for SCD diagnosis varies from 4105 to 13989.

Implementation details

The neural networks were trained using Python 3.6.2 and the TensorFlow package version 1.8.0. The networks were trained and test images were processed on a desktop computer running Windows 10 using an Intel I9-7900X CPU, 64 GB of RAM and one of the computer's two GPUs (NVIDA GTX 1080 Ti). The enhancement network infers each field of view in 0.73 seconds while the classification network inference takes 0.64 seconds per field of view, taking a total of 6.85 sec to process the entire 1.25 mm^2 area of the blood smear. For both of the neural networks, the training image data were augmented by using random rotations and flipping. The hyperparameters and network architecture were chosen specifically for the datasets used in this chapter, adjusted through experimental tuning.

Chapter 5 Virtual multiplexed and microstructured staining of label-free tissues

In addition to performing transformations between images to improve the quality of the images, the same general technique can be used to perform a wide variety of arbitrary transformations where the information required for the transformation is contained within the input to the neural network. Chapter 5 of this dissertation will focus on the transformation from label-free autofluorescence images of thin tissue sections into virtually stained tissues which resemble the same tissue after chemical labeling. Virtual staining of unlabeled tissues has previously been shown to be highly effective at recreating histochemical stains both in quality as well as diagnostically [117]. However, it relied upon tissue-stain-specific neural networks to perform the transformation.

Here, I will demonstrate a new framework that can generate multiple stains using a single deep neural network. This framework can generate virtually stained images using label-free tissue images, in which different stains are merged following a micro-structure map defined by the user. This approach uses a single deep neural network that receives two different sources of information as its input: (1) autofluorescence images of the label-free tissue sample and (2) a "digital staining matrix", which represents the desired microscopic map of the different stains to be virtually generated in the same tissue section. This digital staining matrix is also used to virtually blend existing stains, digitally synthesizing new histological stains. This virtual-staining network was trained and blindly tested using unlabeled kidney tissue sections to generate micro-structured combinations of hematoxylin and eosin (H&E), Jones' silver stain, and Masson's trichrome stain. Using a single network, this approach multiplexes the virtual staining of label-free tissue images with multiple types of stains and paves the way for synthesizing new digital histological stains that

can be created in the same tissue cross section, which is currently not feasible with standard histochemical staining methods.

Parts of this chapter have been published in:

- 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: Science & Applications (Nature Publishing Group) DOI: 10.1038/s41377-020-0315-y (2020).

5.1 Introduction

Histological analysis is used to diagnose a wide variety of diseases. It is considered the gold standard for tissue-based diagnostics, with some well-established versions of common stains, such as haematoxylin and eosin (H&E), having been used for over a hundred years [119]. The histological staining process first requires the slicing of a fixed tissue specimen into sections of 2-10 microns, which are then fixed to microscope slides. Histological staining chemically introduces contrast into tissue sections, which can then be analysed and used to screen for diseases through bright-field microscopic imaging of the stained samples. However, histological staining can be a long and labour-intensive process, particularly in the case of special stains such as Jones' silver stain and Masson's trichrome stain. Therefore, the tissue staining process can increase both the time needed for diagnosis and the associated costs.

A wide variety of stains have been developed over the years to enable the visualization of different target tissue constituents. For example, haematoxylin stains cell nuclei, while Masson's

trichrome stain is used to view connective tissue [120]. These stains have also been chemically mixed to enable the visualization of different biomarkers. An example of this is when periodic acid-Schiff (PAS) and Alcian blue stains are used in conjunction to perform differential staining of glycoproteins [121].

In recent years, various methods have been developed as substitutes for the histochemical staining of samples in an attempt to avoid (1) the specimen-destructive nature of the labelling process, allowing tissue preservation for more advanced analysis; (2) the lengthy and laborious labelling steps, saving time and cost; and (3) unnecessary additional biopsies from the same patient due to tissue depletion. Some of the earliest alternative contrast generation methods utilize various processes related to light-matter interaction, including nonlinear microscopy [122], Raman scattering [123], programmable supercontinuum pulses [124] and reflectance confocal microscopy [125]. However, as pathologists (and, more recently, machine learning algorithms) are mainly trained to perform diagnoses using histologically stained specimens, images generated using alternative contrast mechanisms might require additional training to analyse. Recent efforts have also focused on the development of computational methods of creating bright-field microscopy images that closely resemble the stained versions of the same specimens. For example, digitally generated pseudo-stains have been demonstrated using analytical and statistical learning-based approaches that transform an input pixel (or pixel spectrum) into an RGB output pixel [123,126,127]. Some of these pixel-to-pixel transformation approaches have also used rapid staining methods to provide contrast in cell nuclei [122,127].

Recently, emerging deep learning methods have enabled the development of algorithms for learning accurate transformations between many different imaging modalities [8,12,38,128,129]. Notably, by utilizing the statistical correlations between the structures in images of unstained tissue

slides and the structures in images of the same slides once stained, an unstained tissue sample can be virtually stained by a trained deep neural network without the need for any chemical processing. For example, using deep learning, autofluorescence images of unlabelled tissue samples have been virtually stained with various types of stains [117]. These virtual stains were validated through a blind study involving a team of board-certified pathologists, revealing that there is no statistically significant difference in quality between a virtually stained image and a standard histochemically stained version of the same sample imaged with a bright-field microscope in terms of either stain quality or diagnostic information. Various other techniques for performing virtual staining of unlabelled tissue images have also been demonstrated, for example, by using quantitative phase images [130] or a combination of two-photon excitation and fluorescence lifetime imaging [131]. Researchers have also used deep learning to improve the accuracy of diagnosis using H&E images [132]; it has been shown that deep neural networks can be used to normalize stains, making them more consistent, which allows automated diagnostic analysis to be performed more easily [133].

In this chapter, I demonstrate a novel machine-learning-based framework that allows users to virtually create micro-structured and multiplexed histological stains in the same tissue section using only a single artificial deep neural network. Using this technique, a trained deep neural network can (1) perform virtual staining in a defined region of interest following a micro-structure map defined by the user and (2) achieve the blending of multiple virtual stains and the synthesis of new digital stains. This framework uses the stain type as the input class for a conditional generative adversarial network (GAN) to transform the input images, consisting of pairs of unlabelled autofluorescence images of the same tissue sample, into a virtually stained image of the same label-free sample. To do so, I introduce a “digital staining matrix”, which is used as part of

the input to the deep network, to spatially encode the stain type, i.e., each pixel can be virtually stained using a different stain type or a different set of histological stains (see Figure 5.1).

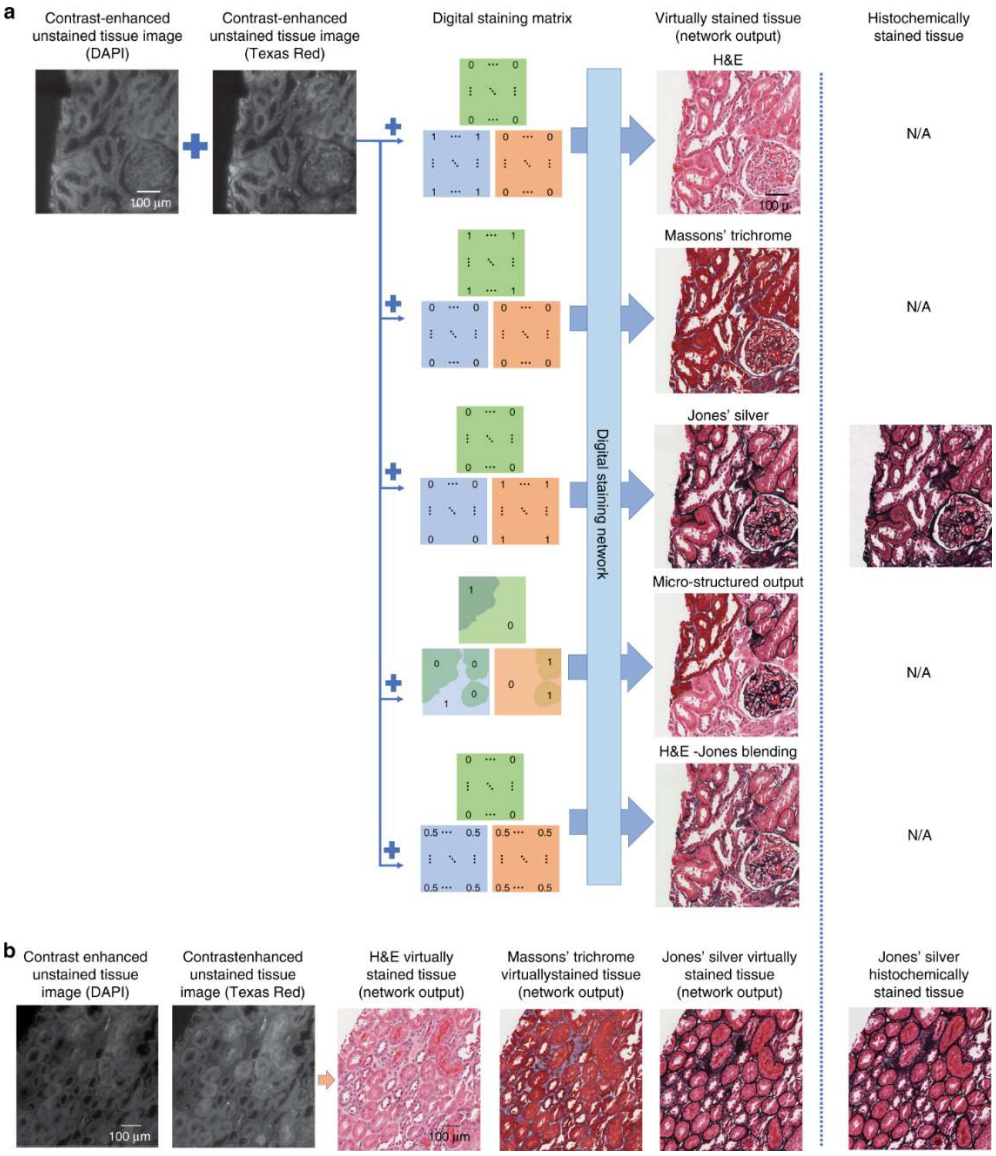

Figure 5.1 Demonstration of multiple stains being virtually generated using a class-conditional neural network and images in two autofluorescence channels (DAPI and Texas Red) of a label-free tissue sample. (a) Steps involved in virtually creating the various stains. By adding a class condition to the network using a digital staining matrix, a single network can be used to generate multiple stains or a blending of stains in the same tissue cross section on demand. (b) A second field of view demonstrating the three digital stains generated using a single trained network. Contrast-enhanced unstained tissue images are provided for visual guidance; unprocessed raw versions of these images were used as the input to the neural network. N/A (not available) refers to the fact that once a tissue section has been histochemically stained with one type of stain, it cannot be subsequently stained with other stains; therefore, the comparison includes N/A entries.

To demonstrate the utility of this technique, a single neural network was trained to virtually stain unlabelled autofluorescence images of kidney needle core biopsy tissue sections with H&E, Jones' silver stain, and Masson's trichrome stain following a user-defined micro-structure map, as illustrated in Figure 5.1. Synthesizing different histological stains and their combinations following a user-defined micro-structure in the same tissue section is currently not feasible with a standard histological staining process, in which different stains are histochemically processed in different tissue sections, leading to tissue depletion. This approach entirely eliminates the need for this, preserving tissue for further analysis while also paving the way for the on-demand synthesis of new digital histological stains in the same tissue section.

5.2 Results

As summarized in Figure 5.1, I demonstrate a method that can be used to perform virtual staining of unlabelled tissue sections using two channels of tissue autofluorescence along with a digital staining matrix, which are used as inputs to a trained deep neural network. The framework was demonstrated using kidney tissue and three different stains, namely, H&E, Masson's trichrome, and Jones' silver stain, as these stains are jointly used for practical renal disease diagnostics. Visualizations of comparisons between histochemically and virtually stained tissue sections can be seen in Figure 5.1 and Figure 5.2.

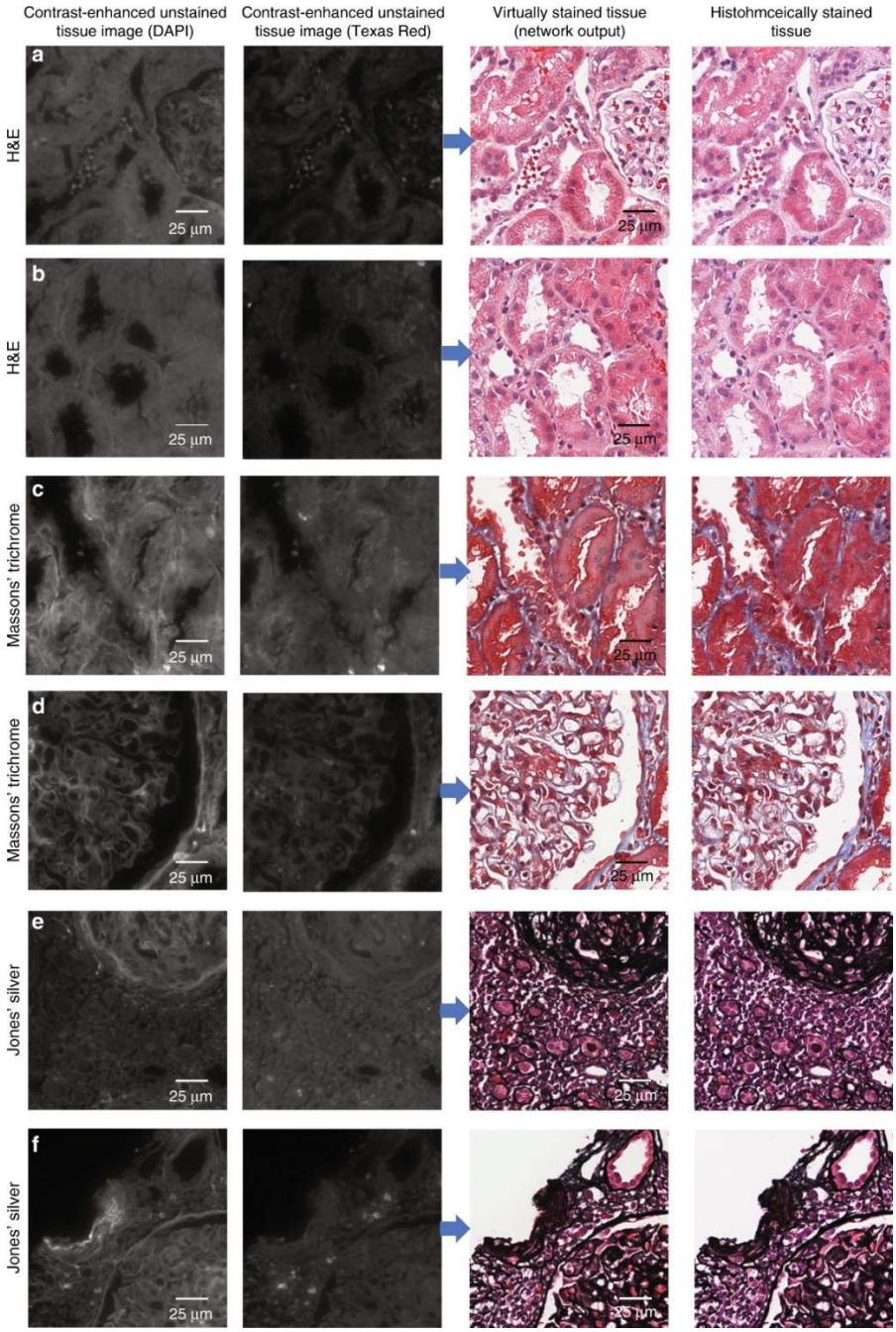

Figure 5.2 Examples of various fields of view that have been virtually stained using the presented multistain network. Co-registered histochemically stained fields of view of the same samples are also shown to the right, and the unstained autofluorescence images are shown to the left to permit direct comparison. (a,b) Tissue stained with H&E, (c,d) tissue stained with Masson's trichrome, and (e, f) tissue stained with Jones' silver. Contrast-enhanced unstained tissue images are provided for visual guidance; unprocessed raw versions of these images were used as the input to the neural network.

The accuracy of the network inference outcomes was validated using quantitative metrics. These quantitative metrics were calculated to confirm that the images generated by the multistain network are highly accurate and that they are equivalent to the images generated using a previously validated single-stain neural network [117]. The first quantitative metric used is the structural similarity index [111] (SSIM), which is defined as:

$$SSIM(a,b) = \frac{(2\mu_a\mu_b + C_1)(2\sigma_{a,b} + C_2)}{(\mu_a{}^2 + \mu_b{}^2 + C_1)(\sigma_a{}^2 + \sigma_b{}^2 + C_2)} \tag{5.1}$$

where μ_a and μ_b are the averages of a and b, the two images being compared; σ_a and σ_b are the standard deviations of a and b; σ_{ab} is the cross-covariance of a and b; and C_1 and C_2 are stabilization constants that are used to avoid division by zero.

Table 5.1 Comparison of SSIM values among the outputs of the different networks and the corresponding histochemically stained tissue images. The averages and standard deviations were calculated across four measured tissue sections.

Stain type	(1) Multistain network output vs. histochemically stained tissue		(2) Single-stain network output vs. histochemically stained tissue		(3) Multistain network output vs. single-stain network output		Total number of image patches compared
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation	
H&E	0.898	0.021	0.905	0.022	0.967	0.006	198
Masson's trichrome	0.850	0.011	0.855	0.023	0.942	0.010	207
Jones' silver	0.803	0.007	0.803	0.010	0.917	0.007	118

Table 5.1 reports the average SSIM values across four unique blindly tested kidney tissue blocks, each from a different patient. Each of these blocks, in turn, was divided into 16 to 60 patches (1224×1224 pixels, or 0.16 mm^2 per patch), each comprising an unlabelled

autofluorescence image pair and its co-registered histochemically stained counterpart (see the Methods section). Because this comparison relies on histochemical staining of the same tissue section, a different section from each tissue block was used for each of the three different stain types. The variation in the number of patches is partially due to the variations in the size of the tissue blocks among patients; furthermore, images that could not be successfully co-registered due to, for example, histochemical-staining-induced tissue distortions were excluded from the SSIM calculations. Three different SSIM values were calculated for each stain type to prove that this new virtual staining technique is successful: (1) the SSIM between the output image of the conditional multistain network and the corresponding image of the histochemically stained tissue, (2) the SSIM between the output of a previously validated [117] single-stain network architecture (see the Methods section) and the corresponding image of the histochemically stained tissue, and (3) the SSIM between the outputs of the multistain network and the single-stain network for each of the three stains. As shown in Table 5.1, a high structural similarity is found for all three cases. Furthermore, the SSIM values calculated for cases (1) and (2) are found to be very similar, indicating that the images generated by the multistain network achieve the same virtual staining performance as was previously reported and validated using the single-stain network [117]. The particularly high structural similarity between the two different virtual staining techniques, i.e., case (3), is also important because the corresponding images are perfectly co-registered since they were generated from the same raw fluorescence images. Together, these results suggest that the presented multistain network generates highly accurate virtually stained images. This demonstrates that the images generated by the multistain network are similar to the corresponding bright-field images of histochemically stained tissue and are of the same quality as those generated by the single-stain network.

These different sets of comparisons between SSIM values are required because the SSIM values between any virtually stained and histochemically stained images depend on a number of factors, some of which are external to the performance of the trained neural network. Perfect co-registration is not feasible, particularly since physical changes are made to the tissue during the actual staining process, [117] somewhat lowering the structural similarity values regardless of the success of the virtual staining network. Furthermore, one of the major benefits of deep-learning-based virtual staining is stain normalization, as the network output will not exhibit the staining variability of the standard histochemical staining process as performed by histotechnologists [117]. While this is certainly a desired feature and will help to improve the consistency of diagnoses, it also lowers the SSIM values due to the histotechnologist-to-histotechnologist variations that are encountered in the ground-truth images.

Table 5.2 Comparison of brightness and chroma differences (using the YCbCr colour space) between (1) multistain network output and histochemically stained tissue, (2) single-stain network output and histochemically stained tissue, and (3) multistain network output

Stain type	Comparison	Y difference (%)		Cb difference (%)		Cr difference (%)		Total number of image patches compared
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation	
H&E	(1) Multistain network vs. histochemically stained tissue	6.62	3.32	0.51	0.18	1.69	1.18	198
	(2) Single-stain network vs. histochemically stained tissue	7.78	3.48	0.87	0.21	2.04	1.41	
	(3) Multistain network vs. single-stain network	1.48	0.12	0.22	0.03	0.72	0.20	
Masson's trichrome	(1) Multistain network vs. histochemically stained tissue	3.85	1.50	1.34	0.87	2.60	1.16	207
	(2) Single-stain network vs. histochemically stained tissue	5.31	1.32	2.09	1.51	3.00	1.44	
	(3) Multistain network vs. single-stain network	1.96	1.70	0.43	0.19	1.35	0.53	
Jones' silver	(1) Multistain network vs. histochemically stained tissue	8.56	2.01	0.82	0.12	2.45	0.69	118
	(2) Single-stain network vs. histochemically stained tissue	9.07	1.93	1.33	0.21	3.15	0.86	
	(3) Multistain network vs. single-stain network	4.32	1.01	0.34	0.11	1.15	0.24	

As another quantitative metric, the average percentage differences in the brightness and chroma components (using the YCbCr colour space) was next compared for the three cases reported in Table 5.1, i.e., (1) multistain network output vs. histochemically stained tissue, (2)

single-stain network output vs. histochemically stained tissue, and (3) multistain network output vs. single-stain network output. As summarized in

Table **5.2**, similar to the case of the SSIM values, the colour differences for cases (1) and (2) are very similar. Case (3) shows particularly small differences, indicating that the two networks (multistain and single-stain) behave very similarly. The change in the brightness (Y) of the multistain network output with respect to the histochemically stained tissue images is relatively low, ranging from 3.84% to 8.57% depending on the stain. The colour distances (Cb and Cr differences) are even smaller (ranging from 0.51% to 2.60% depending on the stain type; see

Table **5.2**), indicating that the multistain network accurately generates the correct colours that represent each stain. Together, these results further demonstrate that the multistain network is capable of accurate virtual staining of unlabelled autofluorescence images of tissue samples and that the output of the multistain network matches the accuracy of the previously validated tissue- and stain-specific neural network [117].

One of the major advantages of using a class-conditional neural network is that it can perform micro-structured virtual staining of tissue sections. Using the presented method, virtual staining of specific areas or structures within tissue sections can be performed by staining different areas of the tissue in accordance with a given micro-structure map. The digital staining matrix, which defines the micro-structure map used to virtually apply different stains for each specified area, can be generated either manually or through the use of a computer algorithm to select structures based on certain diagnostic criteria. An example of virtual stain micro-structuring in accordance with a manually drawn micro-structure map is shown in Figure 5.3. In this example, the marked areas are virtually stained with Masson's trichrome and Jones's silver stains, while the remaining areas not selected are stained with H&E. A co-registered image of the same field of view (FOV) after histochemical H&E staining is also shown for comparison.

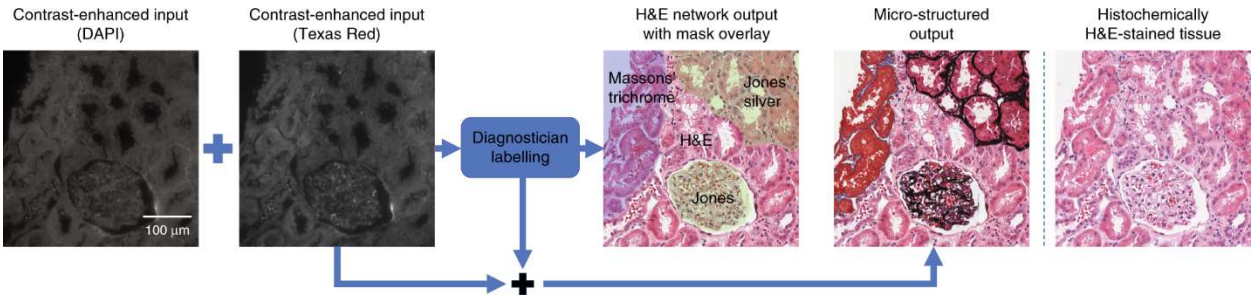

Figure 5.3 Example of multistain micro-structuring. Either a diagnostician or an algorithm can label sub-regions of the unstained tissue, creating on demand a digital staining matrix that defines the microscopic map of multiple stains to be virtually generated in the same tissue section. These labels are used by a single trained network to stain different

areas of the tissue with the desired stains. A co-registered image of the histochemically H&E-stained tissue (same sample) is shown for comparison. Producing a histochemically stained image with the same or a similar microscopic map, with multiple stains in the same tissue section, is not possible with current chemical staining technology.

Stain blending can also be used to digitally synthesize new types of stains. Rather than using the digital staining matrix to generate individual stains, a mixture of multiple stains can be chosen. Such a stain mixture is generated by designing the digital staining matrix to mix two or more stains in the desired tissue areas, simultaneously and at controllable ratios (see Figure 5.4). In other words, the newly generated stain can be tuned on demand by simply changing the ratio between the different values in the digital staining matrix, thus making the different stain combinations more or less pronounced. Figure 5.4 demonstrates several such stain combinations for different pairs of stains. By using these blended stains, aspects of the different stains can be made visible at the same time, which may allow pathologists to more easily view different tissue structures and perform diagnosis. For example, Figure 5.4 (a-e) demonstrates blending between H&E and Jones' silver stain; H&E enables easy differentiation of cell nuclei and cytoplasms [134], while Jones' silver stain provides contrast to basement membranes [135]. By blending these two stains, their associated characteristics can be visualized simultaneously.

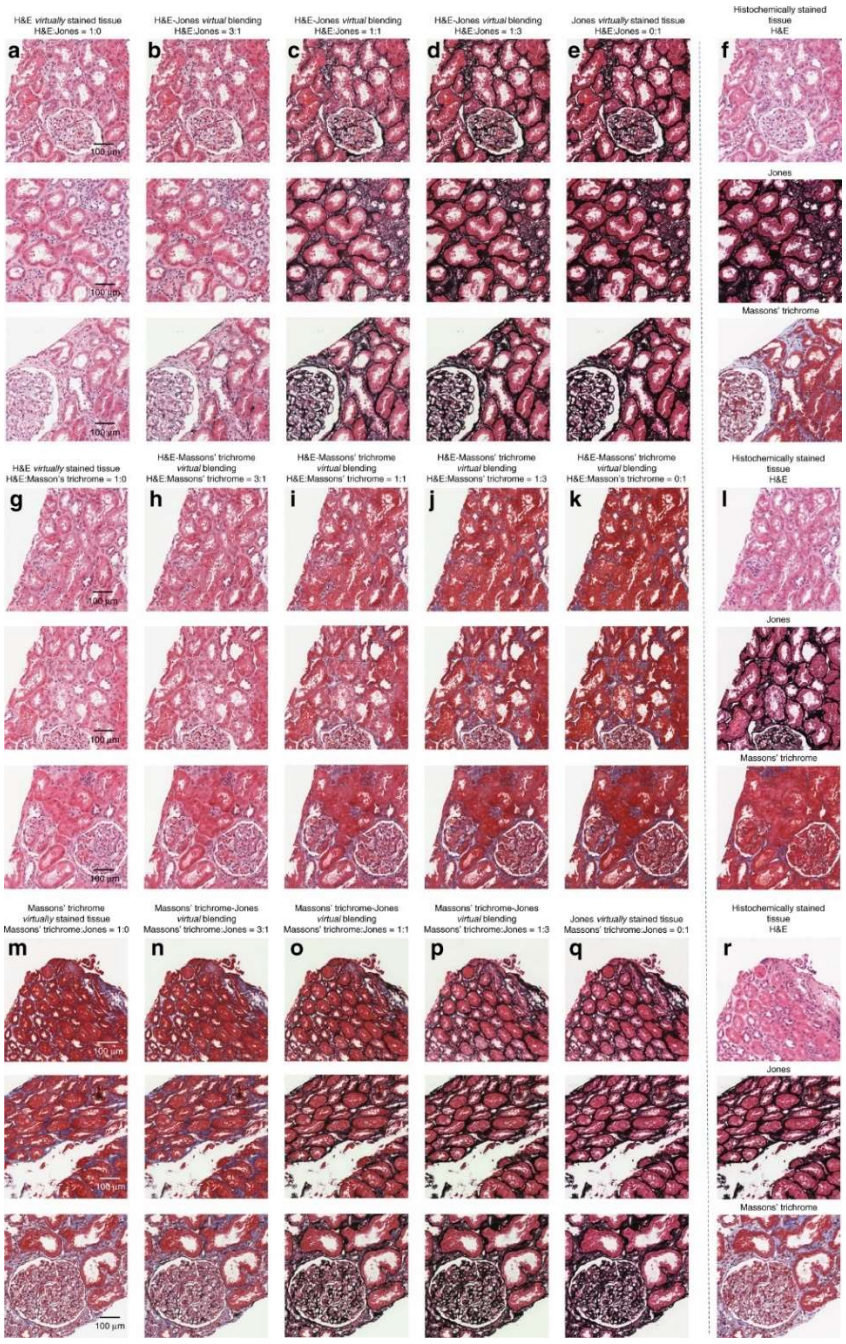

Figure 5.4 Examples of stain blending. (a-e) Kidney tissue that has been virtually stained with varying class-condition ratios of H&E to Jones' silver stain. (g-k) Kidney tissue that has been virtually stained with varying class-condition ratios of H&E to Masson's trichrome stain. (m-q) Kidney tissue that has been virtually stained with varying class-condition ratios of Masson's trichrome stain to Jones' silver stain. (f,l,r) Co-registered images of the histochemically stained tissues (same samples) for comparison (top: H&E; middle: Jones' silver stain; bottom: Masson's trichrome stain).

While the digital stains generated here are completely unique to virtual staining, various histochemical stains have also been mixed together to generate new stain combinations [121]. However, new chemically developed stain combinations can take a large amount of time and resources to reach maturity. In contrast, the stain blending combinations presented here can be developed on demand by simply changing the values of the digital staining matrix until the desired stain is achieved. The virtual stain blending presented here also has a different effect on the tissue than standard histochemical stains, introducing a new mode of micro-structured visualization for pathologists.

5.3 Discussion

In this chapter, I have demonstrated that autofluorescence images of a label-free tissue sample can be used to perform micro-structured and multiplexed virtual staining using a deep neural network. By adding a digital staining matrix to the input of the neural network, it can generate multiple virtual stains upon the same tissue section using a single network. The success of this approach has been validated using kidney tissue sections and three different stains – H&E, Masson's trichrome stain and Jones' silver stain – and allows a pathologist to view the same areas of a sample with all three stains, perfectly matched in the same tissue cross section. The digital staining matrix also allows us to perform micro-structured virtual staining of a label-free sample, in which the sub-area for each stain can be defined either manually or using a separate algorithm. This approach can further be used to perform stain blending by using a digital combination of the stains that the multistain neural network has been trained to apply.

The ability to apply multiple stains to a single tissue section using a single neural network, alongside the newly added capabilities of stain blending, synthesis, and micro-structured virtual staining, has the potential to improve the accuracy and consistency of tissue-based diagnoses.

These new techniques might allow pathologists to obtain more relevant information from tissue than is otherwise possible. By applying stains to specific areas, each tissue constituent can be stained with the most relevant stain. By blending stains, the network is able to simultaneously display information conveyed by each of the separate stains, providing additional channels of information to the pathologists making diagnoses.

These virtual staining techniques also open up opportunities to augment the diagnostic workflow currently used by pathologists and/or machine-learning-based diagnostic algorithms. Virtual staining normalizes the stain quality, improving its consistency and removing variations (caused by, for example, the manual histochemical staining performed by trained professionals) that have not been learned by the neural network [117]. Furthermore, micro-structured staining and stain blending can ensure that the diagnostic platform has access to the most relevant information possible, reducing the amount of unnecessary data viewed/processed by either a pathologist or an algorithm. Consequently, it is believed that the push-pull relationship between the presented virtual staining framework and diagnosticians (human or AI-based) will lead to new uses of the capabilities of this unique framework in pathology and clinical diagnosis, all of which must be clinically validated through rigorous testing and blinded large-scale studies.

5.4 Methods:

Data acquisition

Unstained formalin-fixed and paraffin-embedded (FFPE) kidney tissues were sectioned into thin, 2 μm slices and fixed on standard glass microscope slides. The training and validation dataset for each stain was made up of images obtained from 12 thin tissue sections acquired from unique patients. The test dataset was made up of 4 tissue sections from additional unique patients.

Adjacent tissue sections from each of these patients were used for each of the three stains. Ethical approval for the use of these tissue sections was obtained under UCLA IRB number 18-001029. Using a conventional widefield fluorescence microscope (IX83, Olympus) equipped with a 20×/0.75 NA objective lens (Olympus UPLSAPO) and two separate filter cubes, DAPI (OSFI3-DAPI-5060C, EX 377/50 nm EM 447/60 nm, Semrock) and Texas Red (OSFI3-TXRED-4040C, EX 562/40 nm EM 624/40 nm, Semrock), autofluorescence imaging of these unlabelled tissue sections was performed. The tissue sections were neither deparaffinized nor cover-slipped before being imaged via fluorescence microscopy. The exposure time for the DAPI channel was 50 ms, and that for the Texas Red channel was 300 ms. Once the autofluorescence images had been obtained, the slides were histochemically stained using standard H&E, Jones' silver or Masson's trichrome stain and were then cover-slipped. The staining of the slides was performed by the UCLA Translational Pathology Core Laboratory (TPCL). The histochemically stained slides were then imaged using a scanning microscope (Aperio AT, Leica Biosystems, 20×/0.75NA objective with a 2× adapter) to create the target labels used to train, validate and test the neural network models.

The two unlabelled autofluorescence images of the same tissue sample were used in conjunction with a digital staining matrix to select the stain or set of stains to be generated as the input to a neural network. This input was transformed by a class-conditional generative adversarial network into an equivalent image of a stained tissue section with the same field of view.

Image pre-processing and co-registration

Because the purpose of the deep neural network was to learn the transformation from the unlabelled autofluorescence images of a tissue specimen to an image of a stained specimen, it was crucial that the FOVs were accurately aligned. Furthermore, since more than one autofluorescence

channel was used as the network input, it was necessary to align the different filter channels. To use three different stains (H&E, Masson's trichrome and Jones's silver), the image pre-processing and alignment was implemented for each pair of input and target images from the three staining datasets individually.

The registration steps for matching the autofluorescence and bright-field images followed the process reported by Rivenson et al. [117] One major addition is that when multiple autofluorescence channels (e.g., DAPI and Texas Red) are used as the network input, they must be aligned even if the images in both channels are captured using the same microscope; the corresponding FOVs from the two channels are not precisely aligned at the subpixel level, particularly at the edges of the FOVs. Therefore, an elastic pyramidal registration algorithm was applied to accurately align the multiple autofluorescence channels. This elastic registration algorithm matches the local features of two image channels by hierarchically breaking the image into increasingly smaller blocks and then matching the corresponding blocks [38]. The elastic registration algorithm begins by dividing the image into a grid of 5×5 blocks and calculating block-wise cross-correlations. The distance between the location with the peak correlation and the centre of the block is used to calculate the shift, as the area with the peak correlation is the point with maximum similarity between the two images. By using a weighted average of the translation vector for each block, a 2048×2048-pixel translation map was generated. This translation map was then applied to the Texas Red image to account for the differences between it and the DAPI image. To achieve accurate co-registration, the image was iteratively broken into increasingly smaller blocks until a block size of 100×100 pixels was reached. The final calculated transformation map was then applied to the Texas Red images to ensure that they were aligned with the corresponding images in the DAPI channel. An example of the use of this elastic transformation map can be seen

in Figure 5.5. Finally, the aligned images from both channels were switched to obtain whole-slide images of the samples that contained both the DAPI and Texas Red channels.

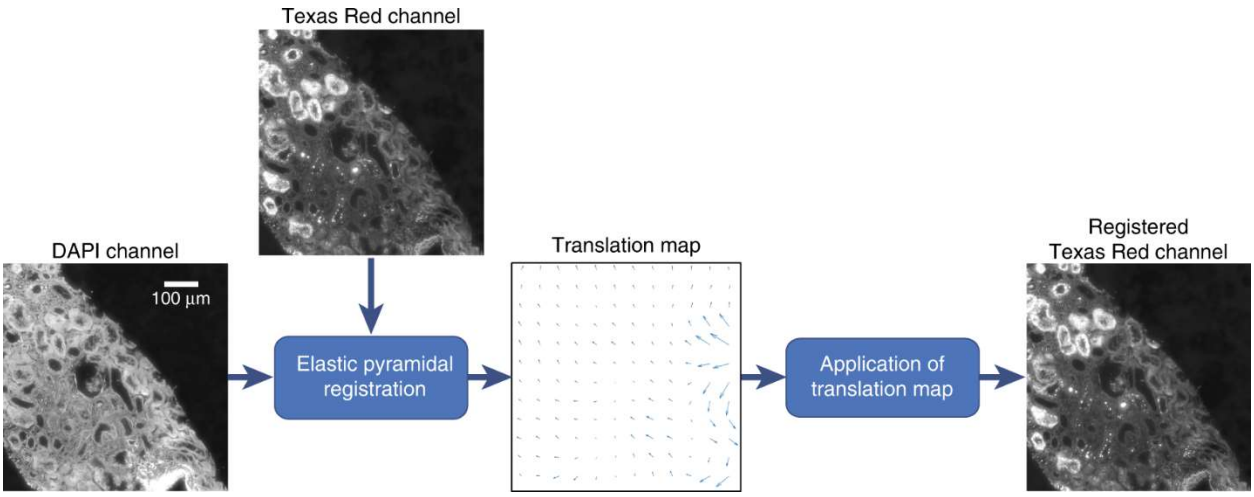

Figure 5.5 Example of co-registration between the DAPI and Texas Red image channels. The translation map was calculated using an elastic pyramidal registration algorithm and then applied to the Texas Red image channel. This process ensured that the two image channels were accurately co-registered with respect to each other.

Co-registration between the fluorescence and bright-field images began with global registration and proceeded with progressive alignment at smaller scales until subpixel-level co-registration was achieved. This first step of this process was to find a rough match by extracting the area of the bright-field image with the highest cross-correlation with a contrast-reversed version of the DAPI image. These images were then further aligned using MATLAB's multimodal image registration feature [116]. Following this registration process, the neural network was trained using patches from these coarsely matched images. Using this network, the autofluorescence images were then virtually stained. However, because only coarse co-registration had been achieved at this point, the accuracy of the virtual staining results could not be satisfactory. Therefore, elastic pyramidal co-registration was then applied to match the histochemically stained images with the initial virtually stained images, resulting in a matched image pair.

Before feeding the aligned pairs into the neural network, we implemented normalization on the whole-slide images generated from the DAPI and Texas Red images. This whole-slide normalization was performed by subtracting the mean value of the entire tissue sample and dividing by the standard deviation of the pixel values (note that background regions were excluded when calculating the mean and standard deviation).

Deep neural network architecture, training and validation

In this study, a class-conditional GAN architecture was used to learn the transformation from the label-free unstained autofluorescence input images to the corresponding bright-field image using three different stains (H&E, Masson's trichrome and Jones' silver). Following the co-registration of the autofluorescence images and the bright-field images, the accurately aligned FOVs were randomly partitioned into overlapping patches of 256×256 pixels and further augmented through rotation and flipping. The patches were then used to train the GAN. During the training process, this class-conditional GAN used a set of one-hot-encoded matrices, together referred to as the digital staining matrix, which was concatenated with the network's 256×256 input image/image stack patches, with each matrix corresponding to a different stain. One way to represent this conditioning is:

$$\tilde{c} = [c_1,\ c_2,\ c_3] \tag{5.2}$$

where $[\cdot]$ denotes concatenation and c_i represents a 256×256 matrix of labels for the i-th stain type (in this example, H&E, Masson's trichrome or Jones' silver). For a pair of input and target images from the i-th stain dataset, c_i was set to be an all-one matrix, while all remaining matrices were assigned values of zero.

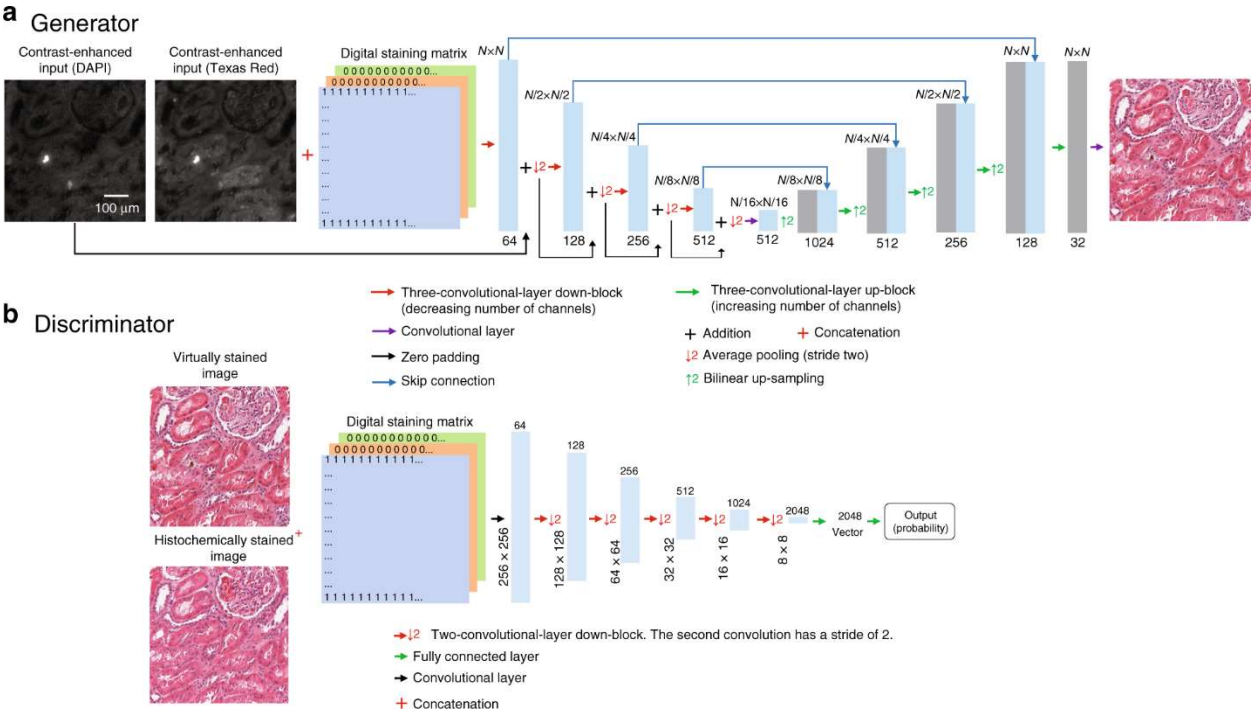

Figure 5.6 Diagram showing the network architecture of the GAN used to perform the transformation. (a) Generator network. (b) Discriminator network.

A GAN is composed of two deep neural networks, a generator and a discriminator (Figure 5.6). During GAN training, the generator learns to perform a statistical transformation to generate a virtually stained image, while the discriminator attempts to distinguish between histochemically stained images and their virtually stained counterparts. The networks improve by learning from one another, improving the quality of the virtually stained images. For this task, the loss functions of the generator and discriminator were defined as:

$$\ell_{generator} = L_1\{z_{label}, G(x_{input}, \tilde{c})\} + \lambda \times TV\{G(x_{input}, \tilde{c})\} + \alpha \times \left(1 - D(G(x_{input}, \tilde{c}), \tilde{c})\right)^2 \quad (5.3)$$

$$\ell_{discriminator} = D(G(x_{input}, \tilde{c}), \tilde{c})^2 + \left(1 - D(z_{label}, \tilde{c})\right)^2$$

where the total variation (TV) operator and mean absolute error (L_1-norm) are used to regularize the generator's output and ensure that it is highly accurate. The TV operator and the L_1-norm are defined as:

$$TV(z) = \sum_p \sum_q \left|z_{p+1,q} - z_{p,q}\right| + \left|z_{p,q+1} - z_{p,q}\right| \tag{5.4}$$

$$L_1(z,\ G) = \frac{1}{P \times Q} \sum_p \sum_q \left|z_{p,q} - G\left(x_{input}, \tilde{c}\right)_{p,q}\right| \tag{5.5}$$

where $D(\cdot)$ and $G(\cdot)$ refer to the outputs of the discriminator and generator networks, respectively; z_{label} denotes the bright-field image of the histochemically stained tissue; and x_{input} represents the input to the neural network. P and Q represent the numbers of vertical and horizontal pixels, respectively, of the image patch, and p and q represent the pixel locations. The regularization parameters (λ and α) were set to 0.02 and 2000, respectively, to accommodate a total variation loss term of approximately 2% of the L_1 loss and a discriminator loss term of 98% of the total generator loss.

For the generator a modified version of the U-net architecture was adopted [4], as visualized in Figure 5.6 (a). This U-net consists of four "down-blocks" followed by four "up-blocks". Each of the down-blocks is made up of three convolutional layers and their activation functions, which together double the number of channels. These convolutional layers are followed by an average pooling layer with a stride and kernel size of two, which effectively down-samples the image. The up-blocks first bilinearly resize the tensors, up-sampling them by a factor of two. This is followed by three convolutional layers and their activation functions. These convolutional layers together reduce the number of channels by a factor of four. Between each of the up- and down-blocks of the same level, a skip connection is used. These skip connections concatenate the output of the

down-blocks with the up-sampled values, allowing data to be passed at each level. Following these down- and up-blocks, a convolutional layer is used to reduce the number of channels to three, which correspond to the three colour channels in the bright-field image.

The discriminator network, visualized in Figure 5.6 (b), receives six input channels. Three channels (YCbCr colour map) come from either the generator output or the target/label, and three come from the one-hot-encoded digital staining matrix. The discriminator architecture contains a convolutional layer that transforms this input into a 64-channel feature map, which is, in turn, passed through a set of five blocks, each consisting of two convolutional layers and their corresponding activation functions. The second of these convolutional layers doubles the number of channels and has a stride of two. These five blocks are followed by two fully connected layers, which reduce the dimensionality to a single channel, which is acted upon by a sigmoid activation function.

The convolutional filter size throughout the GAN is set to 3×3; the outputs of these filters are acted upon by the Leaky ReLU activation function, which is described as:

$$LeakyReLU(x) = \begin{cases} x & for\ x > 0 \\ 0.1x & otherwise \end{cases} \tag{5.6}$$

During training, the learnable parameters were updated using the adaptive moment estimation (Adam) optimizer with learning rates of 1×10^{-4} for the generator network and 2×10^{-6} for the discriminator network. For each step of discriminator training, ten iterations of training were performed for the generator network. The batch size for training was set to 8.

Virtual staining of unlabelled tissue images with a single stain

Once the network had been trained, the one-hot-encoded label $\tilde{c}$ was used to condition the network to generate the desired stained images. In other words, to generate solely the i-th stain, the matrix c_i was set to be an all-one matrix, and the remaining matrices were set to be all zeros.

Stain blending and micro-structured virtual staining of unlabelled tissue images

Following the training process of the neural network model, the conditional matrices can be used in ways different to that in which the model was trained to virtually create new types of stains. The basic encoding rule that should be satisfied can be summarized as follows:

$$\sum_{i=1}^{N_{stains}} c_{i,j,k} = 1 \tag{5.7}$$

In other words, for a given set of indices j and k, the sum over the number of stains on which the network was trained ($N_{stains} = 3$ in this example) should be equal to 1. By modifying the class encoding matrices to use a mixture of multiple classes, the various stains can be blended, creating unique stains with features inherited from the various stains learned by the artificial neural network. Examples of such blended stains are illustrated in Figure 5.4.

Another possible use of the trained multistain neural network is to partition the tissue field of view into different regions of interest (ROIs) and virtually stain each ROI using a different specific stain or blend of a sub-set of these stains:

$$\sum_{i=1}^{N_{stains}} c_{i,j,k} = 1 \quad for \quad j,k \subseteq ROI \tag{5.8}$$

where *ROI* is the defined region of interest in the sample field of view. Multiple non-overlapping *ROI*s can be defined across a field of view, with different stains applied to different regions of

interest or micro-structures. These can be either defined by the user or algorithmically generated. As an example, a user can manually define various tissue areas via a graphical user interface and stain them with different stains. This will result in different tissue constituents being stained differently, as illustrated in Figure 5.1 and Figure 5.3. This ROI-selective staining (micro-structured staining) has been implemented functionality using the Python segmentation package Labelme [136]. Using this package, we can generate logical masks in accordance with labelled ROIs, which are then processed to be the $\tilde{c}_{ROI}$ labels for specific microscopic areas. Other manual, software or hybrid approaches can also be used to implement the selection of certain tissue structures.

Single-stain network used for SSIM calculations

To generate virtually stained images using a single-stain network, a network with the same architecture but excluding the digital staining matrix was used. A separate network was trained for each of the three stains using the portion of the dataset specific to that stain. This single-stain network was implemented followed the approach previously reported [117].

Implementation details

The virtual staining network was implemented using Python version 3.6.0, with the TensorFlow framework version 1.11.0. We implemented the software on a desktop computer with an Intel Xeon W-2195 CPU @2.30 GHz and 256 GB of RAM running the Microsoft Windows 10 operating system. Network training and testing were performed using a single NVIDIA GeForce RTX 2080 Ti GPU. The network was trained for 21000 discriminator training steps over 47 hours. Using a single GPU, inference can be performed at a rate of 3.9 s per 1 mm^2 of unlabelled tissue.

Chapter 6 Transforming H&E stained tissue into special stains

One of the unique aspects of the technology described in Chapter 5 is the ability to generate multiple stains using a single deep neural network. In this chapter, I will describe a method which uses images generated by this technique to create synthetic data, which can in turn be used to train addition neural networks to perform transformations between stains. Since hematoxylin and eosin is the most commonly used stain, used in nearly every case, I will demonstrate the utility of this supervised stain transformation technique by performing a transformation from H&E to special stains (Masson's Trichrome, periodic acid-Schiff and Jones silver stain) using kidney needle core biopsy tissue sections. Based on the evaluation by three renal pathologists, followed by adjudication by a fourth pathologist, I will show that the generation of virtual special stains from existing H&E images improves the diagnosis across several non-neoplastic kidney diseases, sampled from 58 unique subjects (P=0.0095). A second study which found that the quality of the computationally generated special stains was statistically equivalent to those which were histochemically stained will also be presented. This stain-to-stain transformation framework can improve preliminary diagnoses when additional special stains are needed, also providing significant savings in time and cost.

Parts of this chapter have been published in:

- K. de Haan, Y. Zhang, J.E. Zuckerman, T. Liu, A.E. Sisk, M.F.P. Diaz, K. Jen, A. Nobori, S. Liou, S. Zhang, R. Riahi, Y. Rivenson, W.D. Wallace, and A. Ozcan, "Deep learning-based transformation of H&E stained tissues into special stains," Nature Communications DOI: 10.1038/s41467-021-25221-2 (2021)

6.1 Introduction

Histological analysis of stained human tissue samples is the gold standard for evaluation of many diseases, as the fundamental basis of any pathologic evaluation is the examination of histologically stained tissue affixed on a glass slide using either a microscope or a digitized version of the histologic image following the image capture by a whole slide image (WSI) scanner. The histological staining step is a critical part of the pathology workflow and is required to provide contrast and color to tissue by facilitating a chromatic distinction among different tissue constituents. The most common stain (otherwise referred to as the routine stain) is the H&E stain, which is applied to nearly all clinical cases, covering ~80% of all the human tissue staining performed globally [138]. The H&E stain is relatively easy to perform and is widely used across the industry. In addition to H&E, there are a variety of other histological stains with different properties which are used by pathologists to better highlight different tissue constituents. For example, Masson's trichrome (MT) stain is used to view connective tissue [120] and periodic acid-Schiff (PAS) can be used to better scrutinize basement membranes. The black staining in the Jones Methenamine silver stain offers sharp contrast to visualize glomerular architecture and enables the pathologist to recognize subtle basement membrane abnormalities resulting from remodeling due to various forms of injury. These features have importance for certain disease types such as non-neoplastic kidney disease [139]. These non-H&E stains are also called special stains and their use is the standard of care in the pathologic evaluation of certain disease entities including non-neoplastic kidney, liver and lung diseases, among others.

The traditional histopathology workflow can be time consuming, expensive, and requires laboratory infrastructure. Tissue must first be sampled from the patient, fixed either through freezing in Optimal Cutting Temperature (OCT) compound, or paraffin embedding, sliced into thin (2-10 μm) sections, and mounted onto a glass slide. Only then can these sections be stained using the desired chemical staining procedure. Furthermore, if multiple stains are needed, multiple tissue sections are cut, and a separate procedure must be used for each stain. While H&E staining is performed using a streamlined staining procedure, the special stains often require more preparation time, effort and monitoring by a histotechnologist, which increases the cost of the procedure and takes additional time to produce. This can in turn increase the time for diagnosis, especially when a pathologist determines that these additional special stains are needed after the H&E stained tissue has been examined. The tissue sectioning and staining procedure may therefore need to be repeated for each special stain, which is wasteful in terms of resources, materials and might place a burden on both the healthcare system and patients if there is an urgent need for a diagnosis.

Recognizing some of these limitations, different approaches have been developed to improve the histopathology workflow. Histological staining has been reproduced by imaging rapidly labeled tissue sections (usually by a nuclear staining dye) using an alternative contrast mechanism acquired by e.g., non-linear microscopy [122] or ultraviolet tissue surface excitation [140], and digitally transforming the captured images into user-calibrated H&E-like images [141]. These approaches mainly focus on eliminating tissue fixation from the workflow, targeting rapid intraoperative contrast to unfixed specimens. More recently, computational staining techniques known as virtual staining have been developed. Using deep learning, virtual staining has been applied on label-free (i.e., unstained) fixed and glass slide affixed tissue sections using various

modalities such as autofluorescence [117,118], hyperspectral imaging [142], quantitative phase imaging [130], and others [131,143]. Virtual staining of label-free tissue not only has the ability to reduce costs and allow for faster staining, but also allows the user to perform further advanced analysis on the tissue since the destructive additional sectioning and staining process is avoided that can cause the specimen to be depleted leading to e.g., additional/unnecessary biopsies from the patients [144]. Furthermore, virtual staining of label-free tissue enables new capabilities such as the use of multiple virtual stains upon a single tissue section, stain normalization (i.e., standardization), region-of-interest specific digital blending of multiple stains, all of which are challenging or highly impractically with standard histochemical staining workflows [117,118].

An alternative approach that can be used to bypass histochemical tissue staining is to computationally transform the WSI of an already stained tissue into another stain (this will be referred to as stain transformation). This allows users to reduce the number of physical stains required without making any changes to their traditional histopathology workflow, and also carries many of the benefits of the virtual staining techniques such as improving stain consistency and reduction in stain preparation time. Different stain transformations have been demonstrated in the literature, e.g., transformation of H&E into MT [145] or transformation of fibroblast activation protein-cytokeratin (FAP-CK), a duplex immunohistochemistry (IHC) protocol [146], from images of Ki67-CD8 stained slides. Stain transformations have also been used as a tool to improve the effectiveness of image segmentation algorithms [147,148]. However, many of these stain transformation techniques rely upon unsupervised approaches which use distribution matching losses used by techniques such as cycle consistent generative adversarial networks (GANs) – also known as CycleGANs [149]. It has been shown that, when applied to medical imaging, neural networks trained using only these types of distribution matching losses are prone to

hallucinations [150]. Some researchers have been able to avoid the use of these distribution matching losses and unpaired image data by training networks to perform other stain-to-stain transformations. For example, a stain transformation network was trained using image pairs acquired from adjacent tissue sections [151], while another work used image pairs captured by chemically de-staining and then re-staining the same tissue sections [152].

In this chapter, I present a supervised deep learning-based stain transformation framework, outlined in **Figure 6.1**. The training of this technique is based on spatially-registered (i.e., perfectly paired) image datasets which allow the stain transformation network to be trained without relying on unpaired image data and corresponding distribution matching losses. I demonstrate the efficacy of this technique by evaluating kidney tissues with various non-neoplastic diseases. Non-neoplastic kidney disease relies on special stains to provide the standard of care pathologic evaluation. In many clinical practices, H&E stains are available well before the special stains are prepared, and pathologists may provide a preliminary diagnosis to enable the patient's nephrologist to begin any necessary treatment. In a setting when only H&E slides are initially available, the preliminary diagnosis is followed by the final diagnosis made by examining the special stain images, which are often provided the next working day. Using the presented stain transformation technique (**Figure 6.1**) would alleviate the need to wait for the special stains to be available. This is especially useful for some urgent medical conditions such as crescentic glomerulonephritis or transplant rejection where quick and accurate diagnosis followed by rapid initiation of treatment may lead to significant improvements in clinical outcomes.

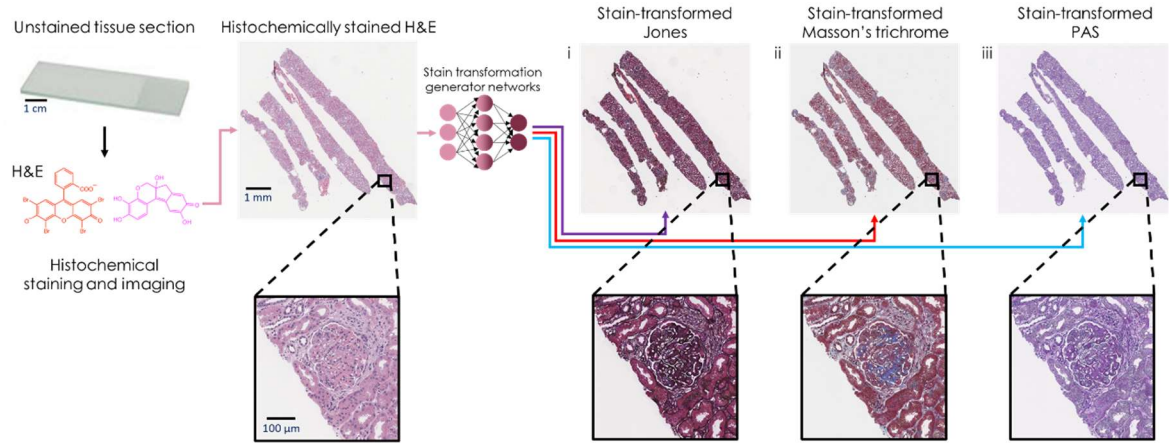

Figure 6.1 Overview of deep learning-based H&E stain transformation into special stains. Histochemical staining of H&E is digitally transformed using a deep neural network into the special stains: (i) generation of JMS (purple arrow); (ii) generation of MT (red arrow); (iii) generation of PAS (blue arrow).

6.2 Results

In order to prove the utility of the stain-transformation technique, I investigated whether it can be used to improve preliminary diagnoses made by pathologists when only H&E is available. To do this, I used stain-to-stain transformation networks to create three additional computationally-generated special stains, i.e., PAS, MT and Jones methenamine silver (JMS), from existing H&E tissue sections. These WSIs were reviewed alongside the existing histochemically-stained H&E images by pathologists (i.e., entirely bypassing the need to stain and wait for new slides). Based on tissue samples from 58 unique patients that are evaluated by three independent renal pathologists (i.e., N=174 total diagnoses), these results revealed that the generation of the three stain-transformed special stains (PAS, MT and JMS) improved the diagnoses in various non-neoplastic kidney diseases. These computationally generated panels of special stains transformed from existing H&E images using deep learning give the pathologists the additional information channels needed for standard of patient care. I show that this unique stain-to-stain transformation workflow can be applied to a variety of diseases, and significantly improves the quality of the

preliminary diagnosis when additional special stains are needed. We believe that this technique has significant utility in enhancing preliminary diagnoses, and could also provide time savings and help to reduce healthcare costs and burden for histopathology labs and patients.

Design and training of stain transformation networks

Deep neural networks were used to perform the transformation between H&E stained tissue and the special stains. To train these networks, a set of additional deep neural networks were used in conjunction with one another. This training workflow relies upon the ability for virtual staining of unlabeled tissue to generate images of different stains using a single unlabeled tissue section (**Figure 6.2**a). By using a single neural network to generate both the H&E images alongside the special stains (PAS, MT, JMS), a perfectly matched training image dataset can be created. However, due to the standardization of the output images generated using the staining network, the virtually stained images (to be used as inputs when training the stain transformation network) must be augmented with additional staining styles to ensure generalization. In other words, network was designed to be able to handle inevitable variability in histochemical H&E staining that is a natural result of (i) differing staining procedures and reagents among histotechnologists and pathology labs, and (ii) differences among digital WSI scanners that are being used. This augmentation is performed by K=8 unique style transfer (staining normalization) networks (**Figure 6.2**b), which ensured that a broad sample space is covered for the presented method to be effective when applied to H&E stained tissue samples regardless of the inter-technician, inter-lab or inter-equipment (e.g., WSI) variations observed at different institutions. Note here that these style transfer networks and the underlying training methods (e.g., CycleGANs) were solely used for H&E stain data augmentation. The use of CycleGANs only expands the sample space of the network inputs during the training, and their outputs were therefore not part of the stain

transformation network loss function. This was possible since perfectly registered training images created by virtual staining of label-free autofluorescence images of tissue were used. This process simultaneously generated both the H&E and special stain images with nanoscopic match in the local coordinates of each virtually stained image pair of the training dataset, which eliminated the need for the use of CycleGANs for stain-to-stain transformation.

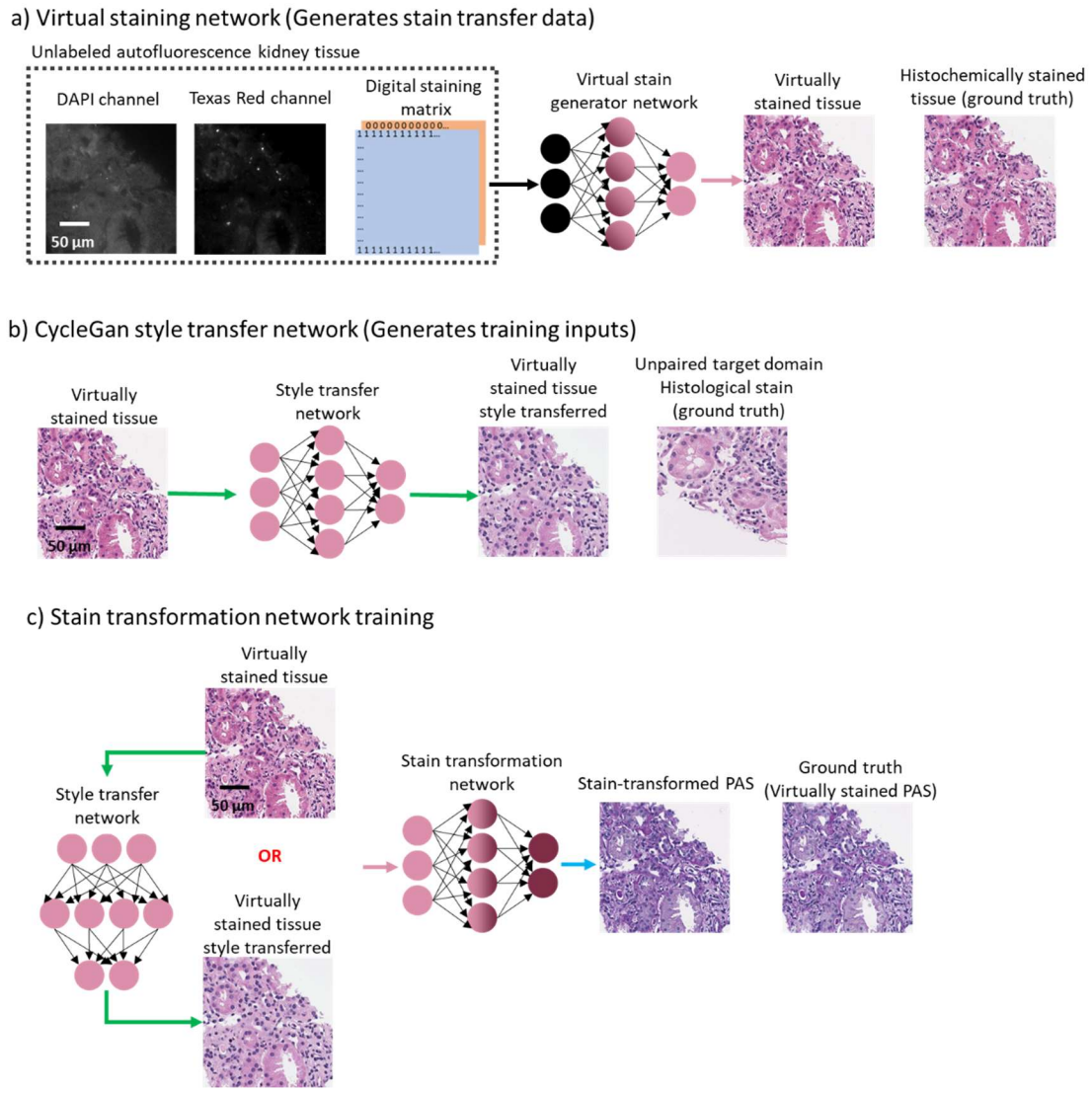

Figure 6.2 Deep neural networks used to generate the training data for the stain transformation network. a) Virtual staining network (pink arrow) which can generate both the H&E and special stain images. b) Style transfer network (green arrow) that is used just to augment the training data. c) Scheme used to train the stain transformation network. During its training, the stain transformation network is randomly given, as the input, either the virtually stained H&E tissue, or an image of the same field of view after passing through one of the 8 style transfer networks. A perfectly matched virtually stained tissue image with the desired special stain (in this example shown: PAS) is used as the ground truth to train this neural network.

Using this image dataset, the stain transformation network is trained, following the scheme shown in **Figure 6.2c.** The network is randomly fed with image patches either coming from the

virtually stained tissue, or the virtually stained images passing through one of the 8 style transfer networks. The corresponding special stain (virtually stained from the same unlabeled field of view) is used as the ground truth regardless of the H&E style transfer. After its training, the network is then blindly tested on a variety of digitized H&E slides taken from UCLA repository, which represent a cohort of diseases and staining variations (all taken from patients that the network was not trained with). The network performs the stain transformation at rate of ~1.5mm^2/s which takes in total ~0.5-1 min for a typical needle core kidney biopsy slide that was used in this study.

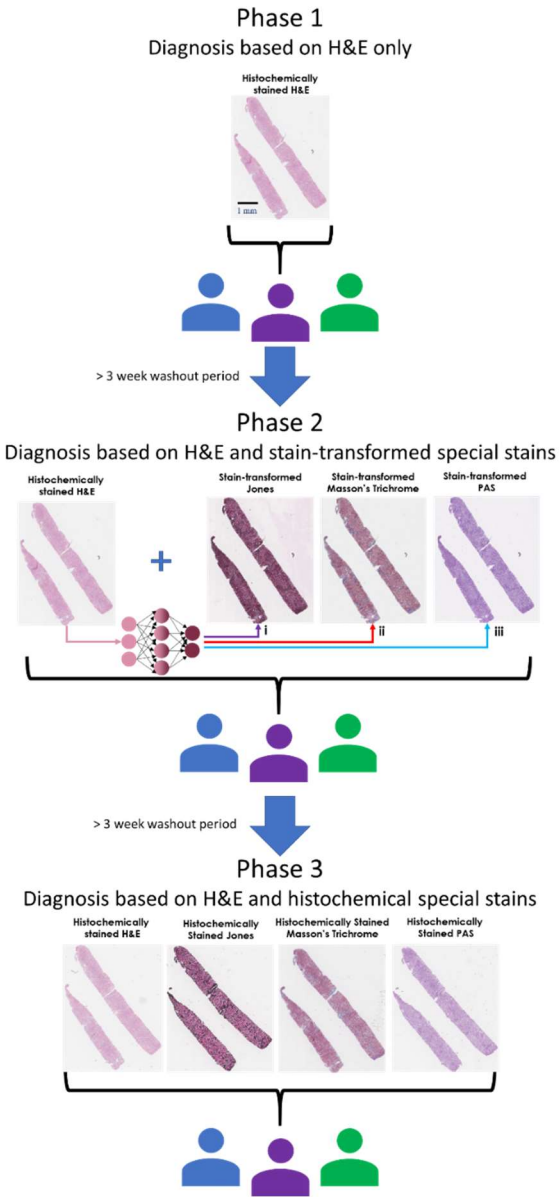

Figure 6.3 Overview of the study design. Phase 1 shows the initial portion of the study where three pathologists review H&E WSIs of N=58 different tissue sections (each from a unique patient). After a >3-week washout period, the second phase of diagnosis is performed, where the same three pathologists view the same WSIs, where, in addition to the H&E, the special stains generated by the stain transformation technique (PAS, Masson's Trichrome, Jones) are provided as well. After an additional >3-week washout period, the third phase of diagnosis is performed, where the same three pathologists again review the same WSIs. For this phase, instead of using special stains generated through the stain transformation technique, the images of all four stains (H&E, PAS, Masson's Trichrome, Jones) come from histochemically stained serial sections. (i) Generation of JMS. (ii) Generation of MT. (iii) Generation of PAS.

Evaluation of stain transformation networks for kidney disease diagnoses

To validate the presented stain transformation technique, a study was performed using WSI data from 58 different H&E stained tissue sections (each corresponding to a unique patient) obtained from an existing database of non-neoplastic kidney diseases. In this blinded study, three board-certified pathologists filled out diagnostic information for each H&E WSI (see the Methods section for details). Following a >3-week washout period, the same pathologists were asked to fill out the same diagnostic information, but along with the H&E, they were also provided the stain-transformed WSIs corresponding to special stains PAS, MT, and JMS, all generated from the existing H&E images. Following a second >3-week washout period, the pathologists were asked to fill out the same diagnostic information. For this third phase, instead of using computationally-generated, stain-transformed special stains, histochemically stained serial tissue sections were given to the pathologists along with the H&E (these sections originated from different depths within the tissue block). A diagram visualizing this study process can be seen in **Figure 6.3**. Following the third round of diagnoses, a fourth board-certified pathologist adjudicated all the results/diagnoses and determined whether the viewing of the neural network generated special stains resulted in an Improvement (I), Concordance (C) or Discordance (D) with respect to the original H&E-only diagnoses. It is important to note that the official reported diagnosis that were used as the ground truth for this study also utilized additional information, such as electron microscopy and immunofluorescence images in order to make these diagnoses.

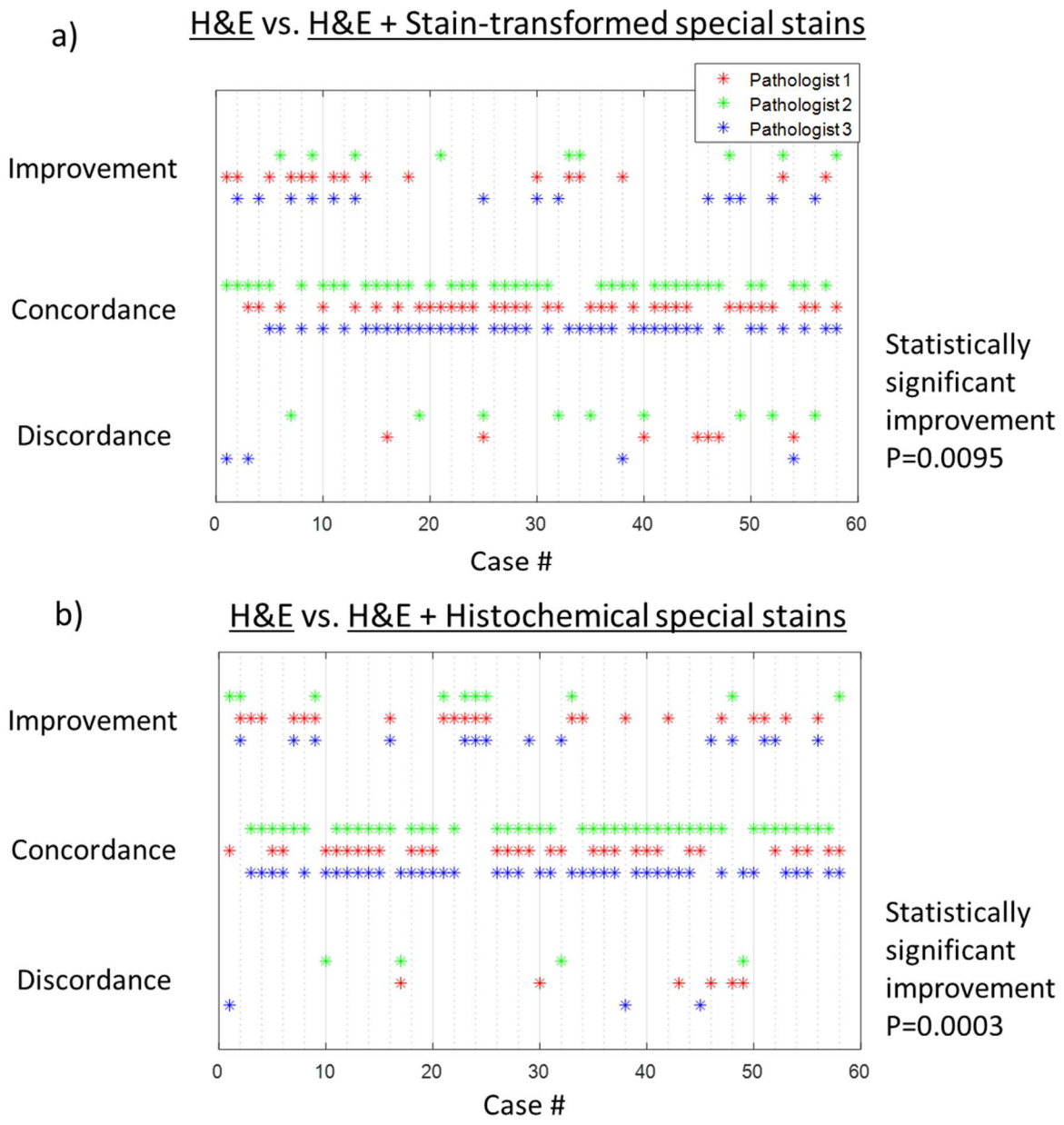

Figure 6.4 Visualization of the improvements, concordances and discordances by case number for the two comparisons. a) Comparison of H&E only vs. H&E and the three stain-transformed special stains coming from the same tissue section. The use of the three stain-transformed special stains results in a statistically significant improvement over H&E only (P=0.0095). b) Comparison of H&E only vs. H&E and the three special stains (all histochemically stained) coming from serial tissue sections. The use of the three histochemically stained special stains results in a statistically significant improvement over H&E only (P=0.0003). P values were calculated using a one-sided t-test. No adjustments for multiple comparisons were needed.

Adjudication of the preliminary diagnoses made by using H&E only and the use of both H&E and stain-transformed special stains across the 58 cases revealed that using stain-to-stain transformations resulted in an average of 13 improved diagnoses (22.4%), 38.3 concordant diagnoses (66.1%) and 6.7 discordant diagnoses (11.5%) across the three pathologists. A total of 10 cases had an improved preliminary diagnosis by 2 or more pathologists, while 3 cases had a discordant diagnosis by more than one pathologists (see **Figure 6.4**). When comparing the diagnoses made with only H&E against those made with H&E alongside the histochemically stained special stains from serial tissue sections, an average of 15 improved diagnoses (25.8%), 38.6 concordant diagnoses (66.6%) and 4.3 discordant diagnoses (7.4%) were found across the three pathologists and 58 cases. For this second comparison, 12 cases were improved by two or more pathologists, while 2 cases were discordant for more than one pathologist (see **Figure 6.4**).

These results show that the additional 3 virtual special stains improve the diagnostic outcome over a single histochemically stained H&E slide for a myriad of non-neoplastic diseases (P=0.0095, using a one-tailed t-test). These stain-to-stain transformation results are also in line with the level of improvement demonstrated when the pathologists had access to the H&E and 3 additional sections that are histochemically stained with the corresponding special stains (P=0.0003, using a one-tailed t-test) over a single histochemically stained H&E slide. In addition to these, a secondary analysis was used to compare differences in the proportion of improvements, concordances and discordances for each of these two comparisons reported in **Figure 6.4**a,b. To do this, three separate chi-square tests were used – one for each pathologist (see the Methods section). These tests found that, while the histochemically stained tissue performed better for all three pathologists, the differences between the two comparisons were not statistically significant (with P values of 0.60, 0.34, and 0.92 for the first, second, and third pathologist, respectively).

For each of the diagnoses marked as improvements, the pathologists were able to provide more accurate characterization or a more complete diagnosis. As an example, **Figure 6.5** demonstrates the improvement using the presented stain transformation technique for a case used for the preliminary evaluation of the technique (diagnosed with Acute Cellular Rejection and Acute Antibody Mediated Rejection), where all three pathologists had the quality of their diagnoses improved. These improvements appear to be based on the clearer definition of the tubular and glomerular basement membranes in the computationally generated special stains. This biopsy contains very pronounced cellular inflammation that is difficult to precisely localize on a standard H&E stain, as H&E does not give clear contrast to structures such as basement membranes. The computationally generated special stains highlight the tubular basement membranes which allowed all 3 pathologists to see the location of the inflammatory cells and give a more precise characterization of the organ rejection process. Another example is case #2, where two pathologists were able to provide a diagnosis of membranous nephropathy only after review of the stain-transformed JMS stain, which is demonstrated in **Figure 6.6**a. In this case, the generated JMS helped the visualization of changes to the basement membrane which are characteristic of membranous nephropathy.

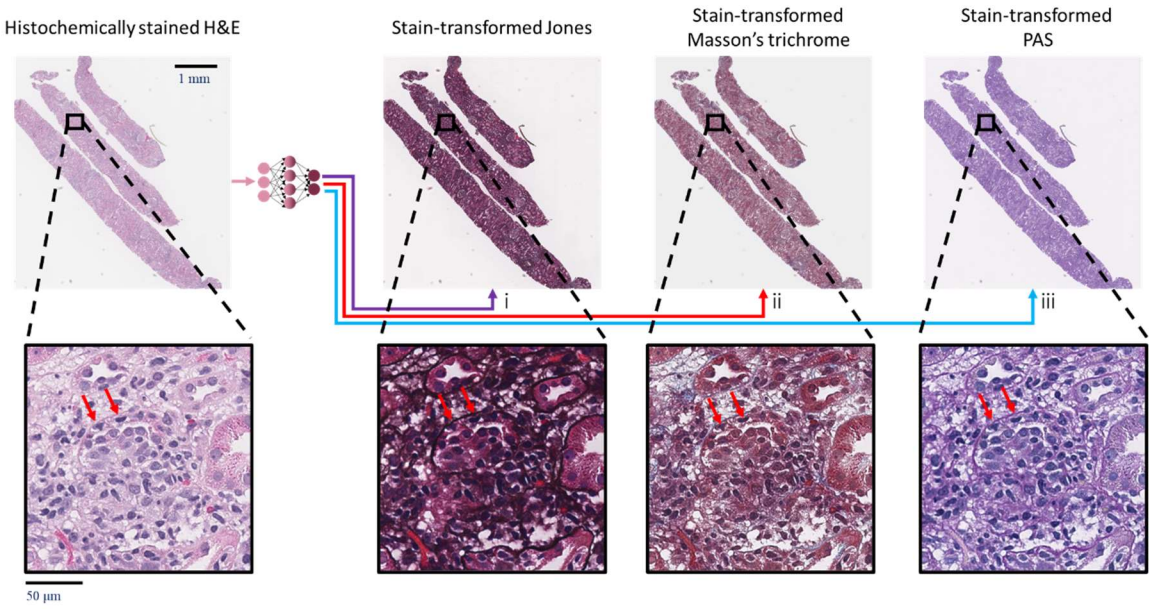

Figure 6.5 Examples of improved diagnoses fostered by the stain-transformed special stains. These WSIs are generated using the stain transformation technique. In this case, the addition of the computationally generated special stains improved all three of the diagnoses made by the pathologists. The red arrows point to a region, where the special stains help highlight inflammatory cells within the tubule, otherwise the boundary of the tubules cannot be seen with the H&E stain only. (i) Generation of JMS. (ii) Generation of MT. (iii) Generation of PAS. A total of 58 cases were viewed by 3 pathologists to perform the statistical analysis.

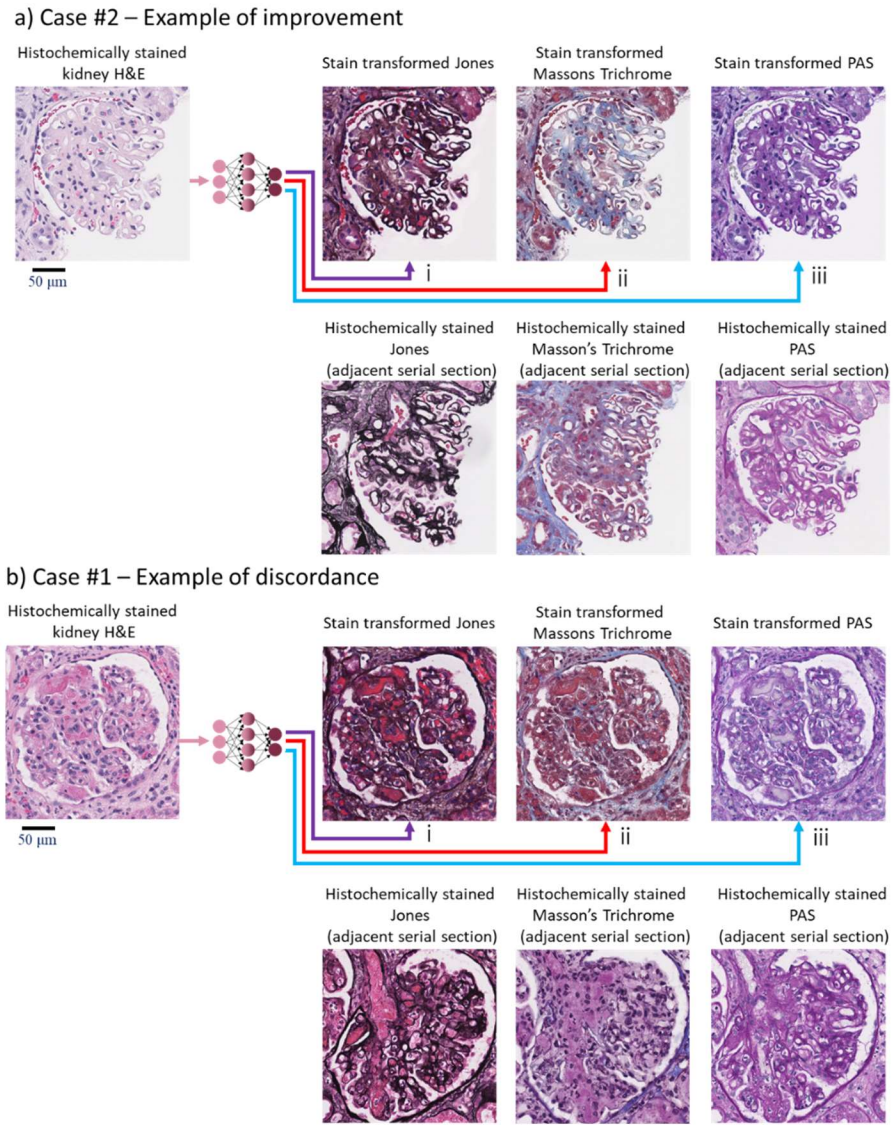

Figure 6.6 Examples of improved and discordant diagnosis achieved by the stain-transformed special stains. a) Example of improved diagnosis fostered by the stain-transformed special stains. For case #2 (in **Figure 6.4**), the basement membrane changes that are characteristic of membranous nephropathy (subepithelial spikes and basement membrane holes) are only appreciated after reviewing the stain-transformed JMS. The bottom images exemplify histochemically stained images of adjacent serial sections of the patient sample; that is why they correspond to different sections within the tissue block. b) Example of the discordance demonstrated between the H&E and computationally generated special stains for case #1 (in **Figure 6.4**). In this field of view, the fibrin thrombi are grey-yellow in color on the stain-transformed PAS stain rather than pink-red. (i) Generation of JMS. (ii) Generation of MT. (iii) Generation of PAS. A total of 58 cases were viewed by 3 pathologists to perform the statistical analysis.

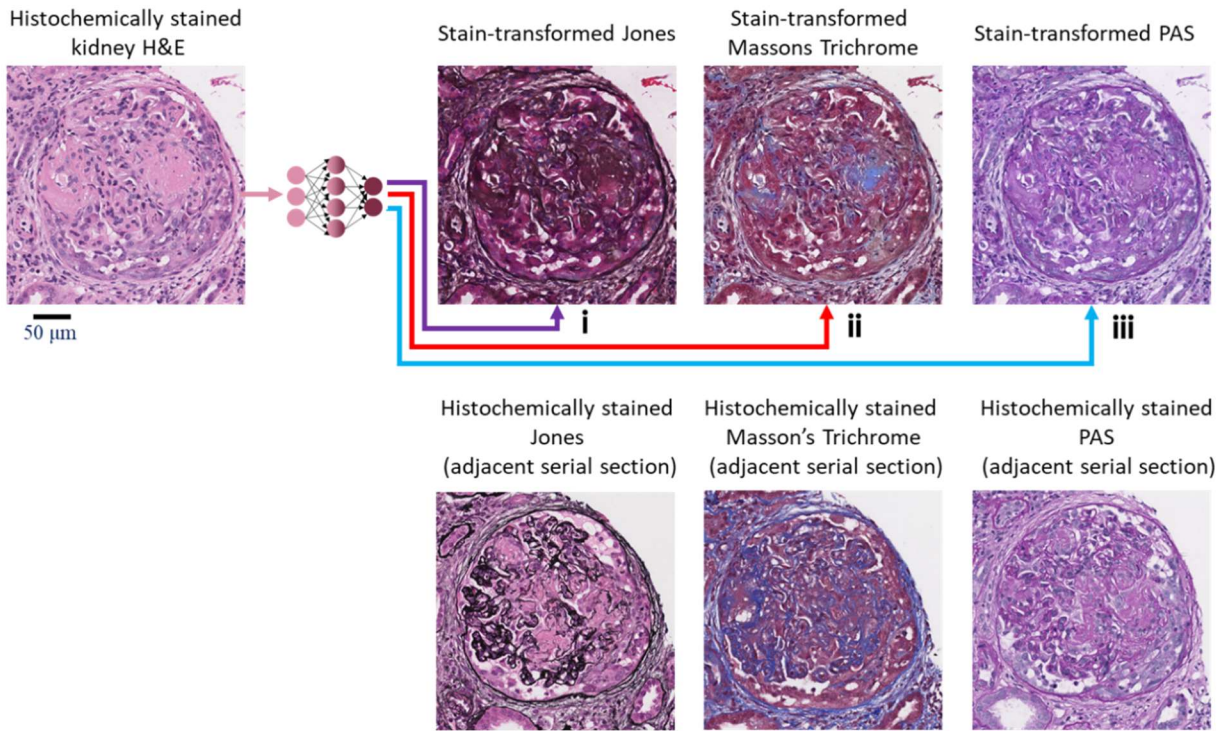

Figure 6.7 Example of the discordance demonstrated between the H&E and stain-transformed special stains for case # 3. The amyloid is darker than normal in the stain-transformed JMS stain for this field of view. (i) Generation of JMS. (ii) Generation of MT. (iii) Generation of PAS. A total of 58 cases were viewed by 3 pathologists to perform the statistical analysis.

The discordances were broken up into two categories: those which were determined to be due to pathologist interpretation error (e.g. case #7), and those which are likely due to mis-representation of the image on the virtual stains. As an example, in case #1, the fibrin thrombi in a case of thrombotic microangiopathy (TMA) appeared too pale on the stain-transformed PAS stain. An example field of view (FOV) with the matching histochemically stained FOV from an adjacent serial tissue section can be seen in **Figure 6.6**b. As a second example, in case #3 (amyloidosis), amyloid deposits were darker on the stain-transformed JMS stain than would be typical in histologically stained slides (an example FOV can be seen in **Figure 6.7**). It is worth emphasizing that in both of these cases (#1 and #3), two of the three pathologists were able to

make concordant diagnoses. Furthermore, one pathologist made a more definitive diagnosis of TMA with the aid of the stain-transformed special stains in case #1 in addition to the original images of the histochemically stained H&E.

We should note that previous research on statistical evaluation of intra-observer decisions revealed a small intra-observer disagreement rate of ~4% when the same cases are viewed by the same pathologist at two different time points [153]. This could potentially account for the discordance in some of the cases such as #7, which was determined to be due to pathologist interpretation error.

Evaluation of the quality of stain-transformed special stain images

An additional study was performed to assess the quality of the stains generated by the stain transformation network. For this study, three pathologists rated the quality of various aspects of the stains generated using the stain transformation network as well as the images of histochemically stained tissue from serial tissue sections. The pathologists each viewed 16 unique rectangular FOVs (with dimensions ranging from ~150 μm ×175 μm to ~375 μm ×500 μm) coming from the three validation slides used during the training of the neural network. These same FOVs were scored for each of the 3 generated special stains, as well as for the same region of the tissue in a serial histochemically stained tissue section. The 16 FOVs were randomly chosen by one of the pathologists to be representative of the tissue sections used, and only in-focus areas where the tissue is unbroken in all of the tissue sections were selected. The FOVs were ordered randomly, and each pathologist rated every image twice – before and after the image randomly being rotated or flipped (resulting in a total of 32 ratings for each stain and image type). Overall, each one of the three pathologists rated 192 FOVs (half virtually stained and half histologically stained).

The pathologists scored four aspects of each FOV on a scale from 1 to 4, where 4 is perfect, 3 is very good, 2 is good enough (passable), and 1 is not acceptable, for a total of 2304 unique assessments/ratings made by three pathologists. The Masson's trichrome stain was rated for overall stain quality, nuclear detail, cytoplasmic detail, and extracellular fibrosis quality. The PAS and Jones Silver stains were rated based on their overall stain quality, along with nuclear detail, cytoplasmic detail and basement membrane detail.

Table 6.1 Quality comparison between the stains generated by the stain transformation network and the histochemically stained tissue, where 4 is Perfect, 3 is Very Good, 2 is Good Enough (Passable), and 1 is Not Acceptable. The scores are the average of 16 fields-of view coming from 3 tissue sections, each rated twice by three separate pathologists.

a) Masson's Trichrome				
	Stain quality score	Nuclear detail	Cytoplasmic detail	Extracellular Fibrosis
Stain transformation	3.19	3.39	3.24	3.11
Histologically stained	3.09	2.95	3.19	3.30
Stain transformation std. error (between pathologists)	0.52	0.35	0.47	0.71
Std. error histological (between pathologists)	0.21	0.27	0.25	0.43
b) PAS				
	Stain quality score	Nuclear detail	Cytoplasmic detail	Basement membrane detail
Stain transformation	3.40	3.53	3.38	3.39
Histologically stained	3.51	3.49	3.41	3.53
Stain transformation std. error (between pathologists)	0.41	0.26	0.39	0.44
Std. error histological (between pathologists)	0.33	0.33	0.42	0.33

c) Jones Silver Stain				
	Stain quality score	Nuclear detail	Cytoplasmic detail	Basement membrane detail
Stain transformation	3.84	3.70	3.70	3.91
Histologically stained	3.88	3.72	3.82	3.98
Stain transformation std. error (between pathologists)	0.13	0.22	0.15	0.05
Std. error histological (between pathologists)	0.06	0.16	0.01	0.02

Table 6.1 shows the mean score for each stain type and quality metric. This table shows that the difference in stain quality for all of the measured aspects of each stain is significantly smaller than the standard error between the ratings. This indicates that the stain-to-stain transformation technique achieves quality of stain equivalent to that of the histochemically stained tissue used as the ground truth.

6.3 Discussion

While different approaches have been explored over the past few years to perform a transformation between two stains, the approach presented here has several unique advantages: (1) it involves less chemical processing applied to tissue, without the need for de-staining and re-staining; and (2) the approach is based on supervised training of the stain transformation network using pairs of perfectly registered training images that are created by label-free virtual staining, which constitutes a precise structural fidelity constraint for the distribution loss that is learned by the discriminator, significantly helping its generalization. Stated differently, no stain-to-stain image aberrations or misalignments exist in this training data due to the fact that the source of information (autofluorescence of the label-free tissue) is common for all the virtually stained images. This feature significantly improves the reliability and accuracy of the stain-to-stain

transformation that is learned using this method. These important advantages are enabled by using autofluorescence-based virtual staining of label-free tissue sections with multiple stains to create perfectly paired training image datasets. While in this chapter I showed that autofluoresence can be used to generate contrast from label-free tissue, other contrast mechanisms such as quantitative phase imaging, multi-photon-microscopy, fluorescence lifetime imaging and photo-acoustic microscopy, among others, can also support this supervised training of the presented stain transformation method. The resulting networks that are trained with this methodology can digitally transform any existing chemically-stained tissue image into new types of stains.

Similar to the validation of digital pathology systems in general, a perfect stain-to-stain transformation is not required; the standard of practice is to demonstrate lack of inferiority, which is what we have endeavored to do. Substituting chemically stained slides for stain-transformed slides leads to several advantages, including e.g., decreased slide preparation time, decreased laboratory costs and preservation of tissue for subsequent analysis, if necessary. In the future, the use of computationally generated special stains may make it possible to selectively omit the need for performing actual special stains and save time and laboratory expenses in some settings.

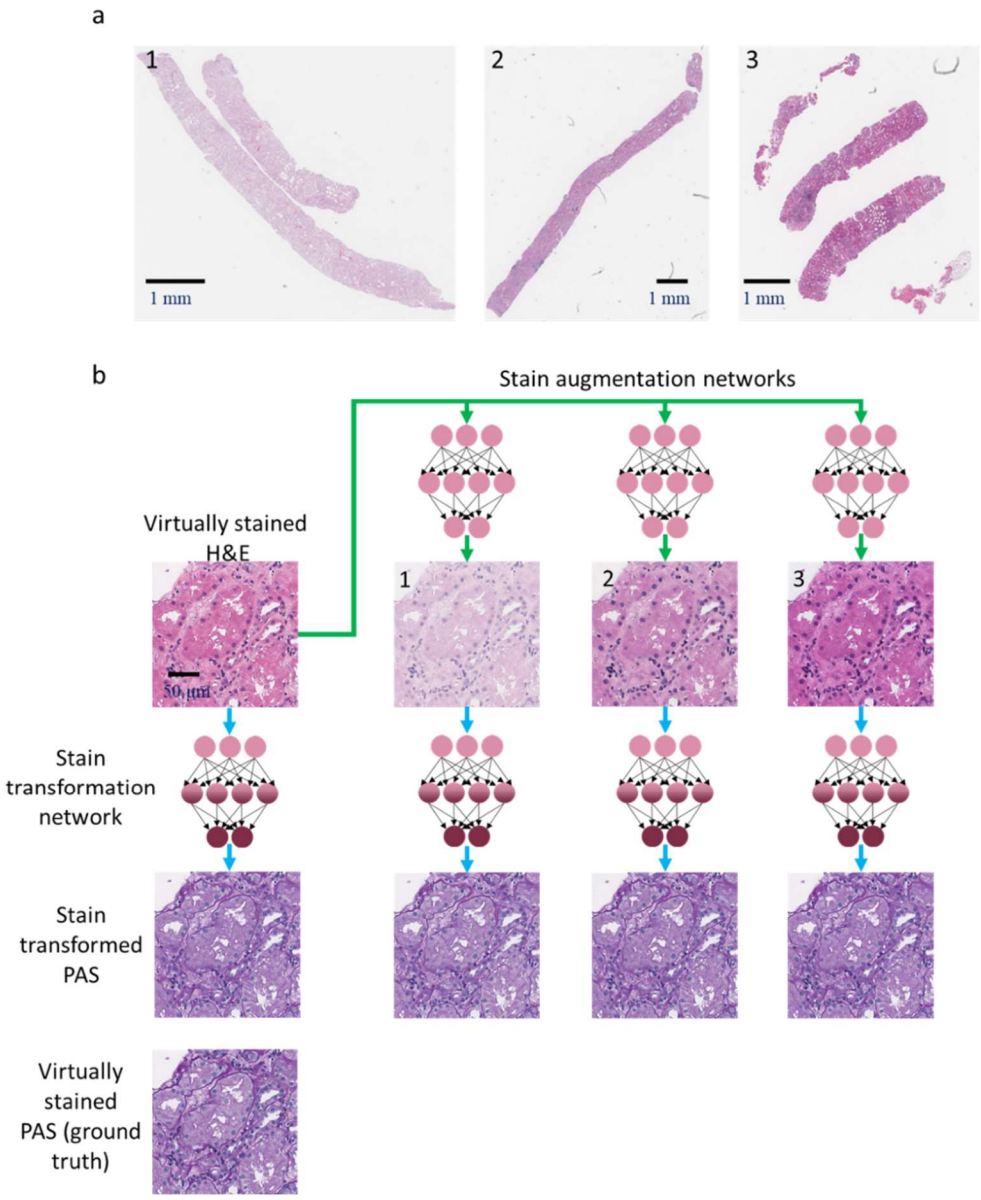

Figure 6.8 Demonstration of the use of style transfer-based image data augmentation. a) Three H&E WSIs that are used as the ground truth distributions for three different stain augmentation networks. b) Demonstration of the effectiveness of the stain transformation technique across different H&E distributions. The three stain augmentation networks (green arrows) perform a style transfer to make the H&E field of view match the distribution of the corresponding three distributions shown in part a. This demonstrates that the stain transformation network (blue arrows) is robust and able to consistently perform accurate transformations when applied to various distributions. A total of 8 style augmentation networks corresponding to 8 unique H&E slides were used to train the stain transformation network.

The ability of this stain-to-stain transformation network to generalize across stain variations is also highly beneficial as there are significant differences among stains produced by different labs and even across stains performed by the same histotechnician (e.g., **Figure 6.8**a demonstrates three examples of such variations for stains produced by the same lab). However, in order a stain transformation technique to be effective for any practical application, the network must generalize across this wide sample space. As one of the key features of virtual staining is stain normalization [117], the network requires data augmentation to better facilitate the learning across a wide input staining distribution. For this purpose, I used a set of 8 CycleGAN networks to perform this stain data augmentation of the H&E dataset used to train the stain transformation network. The use of CycleGAN networks to perform a stain normalizing style transfer has been shown to be more effective than traditional stain normalization algorithms [154]. Furthermore, they have proven to be highly effective at performing data augmentation for medical imaging [155]. By applying these CycleGAN augmentation networks to the training image dataset, I was able to successfully generalize to various slides used for blind testing. Three examples of this CycleGAN-based stain augmentation results are reported in **Figure 6.8**b, which demonstrates that the three different networks are capable of converting the virtually stained tissue to have H&E distributions which match the distributions seen in **Figure 6.8**a. Furthermore, the results show that the same stain transformation network is consistent across these various distributions as there is little variation among the virtual PAS outputs (**Figure 6.8**b). These style normalization/transfer networks used in data augmentation can be easily further expanded upon, if needed, using existing databases of H&E images.

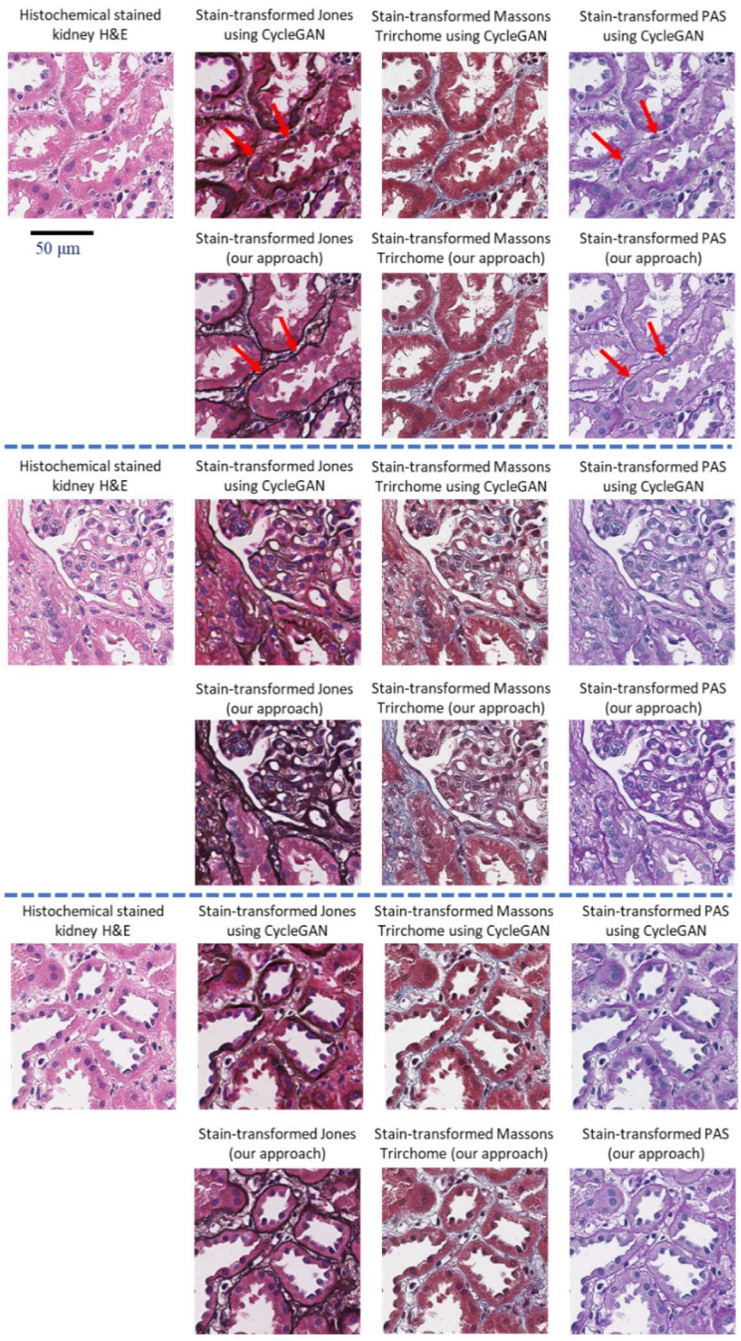

Figure 6.9 Comparison between the performance of the stain-transformation network presented in this work and stain transformations performed by a CycleGAN. For the Masson's Trichome stain there are only minor differences between the quality of the CycleGAN and our method. However, the other two stains are much more difficult for a CycleGAN to perform, and the CycleGAN technique hallucinates features throughout all of the images generated for

these stains. For example, the red arrows point to locations where the basement membrane has been incorrectly labeled by the CycleGAN for the Jones and PAS stains.

As I emphasized earlier, these style transfer networks were only used for H&E stain data augmentation and were not included in the stain transformation loss function. I utilized perfectly registered training images generated by virtual staining of label-free tissue; as a result of this, potential hallucinations or artifacts related to unsupervised training with CycleGANs and unpaired training data are eliminated (as can be seen in **Figure 6.9**). When the same CycleGAN architecture used for the data augmentation is applied to the various stain transformations, a number of clear hallucinations occur. These hallucinations are particularly evident for the PAS and Jones Silver stain, where the networks incorrectly label the tubular basement membranes (see **Figure 6.9**). The tubules are composed of epithelial cells lining basement membranes that stain black on the Jones stain and magenta on the PAS stain. The brush border lining the luminal surface of the epithelial cells is also normally lightly stained black and magenta by the Jones and PAS stains, respectively. The CycleGAN method incorrectly recognized the basement membranes and tubular brush borders leading to incorrect image generation, which is a very significant error. In contrast, the quality and features of the Masson's Trichrome stain appear to be more similar between the two techniques. This is believed to be due to the Masson's Trichrome stain being relatively similar to H&E, while the other stains require significant structural changes which can cause the hallucinations for CycleGANs. These results and observations highlight the significant advantages of the stain-to-stain transformation network compared to standard CycleGAN-based methods.

It is important to note that the current stain-to-stain network is trained to work with H&E stains performed at a few institutions and imaged by different microscopes from the same vendor/model (Leica Biosystems Aperio AT2 slide scanner). Additional data would be required

for the network to generalize to samples imaged using microscopes with different specifications or vendors, or any H&E stains which are performed in a significantly different manner. Furthermore, while this study covers a broad range of diseases, it is still a proof-of-concept. Future studies should be performed which contain both larger training and test datasets in order to conclusively show the technique may be suitable for diagnostic use. Future work may also apply this technique that I have presented to other biomarkers that are currently labeled with IHC to help target specific conditions.

In addition to histological stains, immunofluorescence and electron microscopy [156] based evaluation play significant roles in the standard of care for non-neoplastic kidney biopsy evaluation. In this study we have attempted to isolate the role of standard light microscopy in the non-neoplastic kidney disease evaluation and therefore these other modalities were not included. However, their application in clinical cases would only serve to support the pathologic final diagnosis and add a layer of further confirmation and safety to this resource-saving stain transformation technique.

In this work, I focused on image transformations from H&E to special stains, since H&E is used as the bulk of the staining procedures, covering approximately 80% of all the human tissue staining procedures [138]. However, other stain-to-stain transformations can also be considered. For example, transformations from special stains to H&E or from immunofluorescence to H&E or special stains could be performed using the presented method. This approach allows pathologists to visualize different tissue constituents without waiting for additional slides to be stained with special stains, and I demonstrated it to be effective for clinical diagnosis of multiple renal diseases. Another advantage of the presented technique is that it can rapidly perform the stain transformation

(at a rate of 1.5 mm^2/s on a consumer-grade desktop computer with two GPUs), while saving labor, time, chemicals and can significantly benefit the patient as well as the healthcare system.

6.4 Methods

Training of stain transformation network

All of the stain transformation networks and virtual staining networks used in this dissertation were trained using GANs. Each of these GANs consists of a generator (G) and a discriminator (D). The generator is used to perform the transformation of the input images (x_{input}), while the discriminator network is used to help train the network to generate images which match the distribution of the ground truth stained images. It does this by trying to discriminate between the generated images ($G(x_{\text{input}})$) and the ground truth images (z_{label}). The generator is in turn taught to generate images which cannot be classified correctly by the discriminator. This GAN loss is used in conjunction with two additional losses: a mean absolute error (L_1) loss, and a total variation (TV) loss. The L_1 loss is used to ensure that the transformations are performed accurately in space and color, while the TV loss is used as a regularizer, and reduces noise created by the GAN loss. Together, the overall loss function is described as:

$$l_{\text{generator}} = L_1\{z_{\text{label}}, G(x_{\text{input}})\} + \alpha \times \text{TV}\{G(x_{\text{input}})\} + \beta \times \left(1 - D\left(G(x_{\text{input}})\right)\right)^2 \qquad (6.1)$$

where α and β are constants used to balance the various terms of the loss function. The stain transformation networks are tuned such that the L_1 loss makes up ~1% of the overall loss, the TV loss makes up only ~0.03% of the overall loss, and the discriminator loss makes up the remaining ~99% of the loss (relative ratios change over the course of the training). The L_1 portion of the loss can be written as:

$$L_1(z,\ G) = \frac{1}{P \times Q} \sum_p \sum_q \left| z_{p,q} - G\left(x_{\text{input}}\right)_{p,q} \right| \tag{6.2}$$

where p and q are the pixel indices and P and Q are the total number of pixels in each image. The total variation loss is defined as:

$$\text{TV}\left(G(x_{\text{input}})\right) = \sum_p \sum_q |G(x_{\text{input}})_{p+1,q} - G(x_{\text{input}})_{p,q}| + |G(x_{\text{input}})_{p,q+1} - G(x_{\text{input}})_{p,q}| \tag{6.3}$$

The discriminator network has a separate loss function which is defined as:

$$l_{\text{discriminator}} = D\left(G\left(x_{\text{input}}\right)\right)^2 + \left(1 - D(z_{\text{label}})\right)^2 \tag{6.4}$$

A modified U-net[1] neural network architecture was used for the generator, while the discriminator used a VGG-style[2] network. The U-net architecture uses a set of 4 up-blocks and 4 down-blocks, each containing three convolutional layers with a 3×3 kernel size, activated upon by the LeakyReLU activation function which is described as:

$$\text{LeakyReLU}(x) = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \tag{6.5}$$

The first down block increases the number of channels to 32, while the rest each increase the number of channels by a factor of two. Each of these down-blocks ends with an average pooling layer which has both a stride and a kernel size of two. The up-blocks begin with a bicubic up-sampling prior to the application of the convolutional layers. Between each of the blocks of a certain layer, a skip connection is used to pass data through the network without needing to go through all the blocks. After the final up-block, a convolutional layer maps back to three channels.

The discriminator is made up of five blocks. These blocks contain two convolutional layers and LeakyReLU pairs, which together increase the number of channels by a factor of two. These are followed by an average pooling layer with a stride of two. After the five blocks, two fully connected layers reduce the output dimensionality to a single value, which in turn is input into a sigmoid activation function to calculate the probability that the input to the discriminator network is real, i.e., not generated.

Both the generator and discriminator were trained using the adaptive moment estimation (Adam) [50] optimizer to update the learnable parameters. A learning rate of 1×10^{-5} was used for the discriminator network while a rate of 1×10^{-4} was used for the generator network. For each iteration of the discriminator training, the generator network is trained for seven iterations. This ratio reduces by one every 4000 iterations of the discriminator to a minimum of one discriminator iteration for every three generator iterations. The network was trained for 50000 iterations of the discriminator, with the model being saved every 1000 iterations. The best generator model was chosen manually from these saved models by visually comparing different models. For all three of the generator networks (MT, PAS and JMS), the 15000^{th} iteration of the discriminator was chosen as the optimal model.

The stain transformation networks were trained using pairs of 256×256-pixel image patches generated by the class conditional virtual staining network (label-free), downsampled by a factor of 2 (to match 20× magnification). These patches were randomly cropped from one of 1013 712×712-pixel images coming from 10 unique tissue sections, leading to ~7,836 unique patches usable for training. 76 additional images coming from three unique tissue sections were used to validate the network. These images were augmented using the eight stain augmentation networks, and further augmented through random rotation and flipping of the images. Each of the three stain

transformation networks (MT, PAS and JMS) were trained using images generated by the label-free virtual staining networks from the same input autofluorescence images. Furthermore, the images were converted to the YCbCr color space [157] before being used as either the input or ground truth for the neural networks.

As this stain transformation neural network performs an image-to-image transformation, it learns to transform specific structures using the ~513 million pixels in the dataset that are independently accounted for in the loss function. Furthermore, since the network learns to convert structures which are common throughout many different types of samples, it can be applied to tissues with diseases that the network was not trained with. When used in conjunction with the 8 data augmentation networks which convert the values of these pixels, as well as random rotation and flipping (for an additional 8×) augmentation, there are effectively many billions of pixels which are used to learn the desired stain-to-stain transformation. Because of these advantages, a much smaller number of training samples from unique patients can be used than would be required for a typical classification neural network.

Image data acquisition

All of the neural networks were trained using data obtained by microscopic imaging of thin tissue sections coming from needle core kidney biopsies. Unlabeled tissue sections were obtained from the UCLA Translational Pathology Core Laboratory (TPCL) under UCLA IRB 18-001029, from existing specimen. The autofluorescence images were captured using an Olympus IX-83 microscope (controlled with the MetaMorph microscope automation software, version 7.10.161), using a DAPI filter cube (Semrock OSFI3-DAPI5060C, EX 377/50 nm EM 447/60 nm) as well as a Texas Red filter cube (Semrock OSFI3-TXRED-4040C, EX 562/40 nm EM 624/40 nm) to generate the second autofluorescence image channel.

In order to create the training dataset for the virtual staining network, pairs of matched unlabeled autofluorescence images and brightfield images of the histochemical stained tissue were obtained. H&E, MT and PAS histochemical staining were performed by the Tissue Technology Shared Resource at UC San Diego Moores Cancer Center. The JMS staining was performed by the Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA. These stained slides were digitally scanned using a brightfield scanning microscope (Leica Biosystems Aperio AT2 slide, using 40x/0.75NA objective). All the slides and digitized slide images were prepared from existing specimen. Therefore, this work did not interfere with standard practices of care or sample collection procedures. The H&E image dataset used for in the study came from the existing UCLA pathology database containing WSIs of stained kidney needle-core biopsies, under UCLA IRB 18-001029. These slides were similarly imaged using Aperio AT2 slide scanning microscopes.

Image co-registration

To train label-free virtual staining networks, the autofluorescence images of unlabeled tissue were co-registered to brightfield images of the same tissue after it had been histochemically stained. This image co-registration was done through a multi-step process [41], beginning with a coarse matching which was progressively improved until subpixel level accuracy is achieved. The registration process first used a cross-correlation based method to extract the most similar portions of the two images. Next, the matching was improved using multimodal image registration [116]. This registration step applied an affine transformation to the images of the histochemically stained tissue to correct for any changes in size or rotations. To achieve pixel-level co-registration accuracy, an elastic registration algorithm was then applied. However, this relies upon a local correlation-based matching. Therefore, to ensure that this matching could be accurately performed,

an initial rough virtual staining network is applied to the autofluorescence images [117,118]. These roughly stained images were then co-registered to the brightfield images of the stained tissue using a correlation-based elastic pyramidal co-registration algorithm [69].

Once the image co-registration is complete, the autofluorescence images were normalized by subtracting the average pixel value of the tissue area for the WSI and subsequently dividing it by the standard deviation of the pixel values in the tissue area.

Class conditional virtual staining of label-free tissue

A class conditional GAN was used to generate both the input and the ground truth images to be used during the training of the presented stain transformation networks (**Figure 6.2**a). This class conditional GAN allows multiple stains to be created simultaneously using a *single* deep neural network [118]. To ensure that the features of the virtually stained images are highly consistent between stains, a single network must be used to generate the stain transformation network input (virtual H&E) and the corresponding ground truth images (virtual special stains) that are automatically registered to each other as the information source is the same image. This is only required for the training of the stain transformation neural networks and is rather beneficial as it allows both the H&E and special stains to be perfectly matched. Furthermore, an alternative image dataset made up of co-registered virtually stained and histochemically stained fields of view will present limitations due to imperfect co-registration and deformities caused by the staining process. These are eliminated by using a single class conditional GAN to generate both the input and the ground truth images.

This network uses the same general architecture as the network described in the previous section, with the addition of a Digital Staining Matrix concatenated to the network input for both

the generator and discriminator [118]. This staining matrix defines the stain coordinates within a given image field of view. Therefore, the loss functions for the generator and discriminator are:

$$l_{\text{generator}} = L_1\{z_{\text{label}}, G(x_{\text{input}}, \tilde{\mathbf{c}})\} + \alpha \times \text{TV}\{G(x_{\text{input}}, \tilde{\mathbf{c}})\} + \beta \times \left(1 - D(G(x_{\text{input}}, \tilde{\mathbf{c}}), \tilde{\mathbf{c}})\right)^2 \quad (6.6)$$

$$l_{\text{discriminator}} = D(G(x_{\text{input}}, \tilde{\mathbf{c}}), \tilde{\mathbf{c}})^2 + \left(1 - D(z_{\text{label}}, \tilde{\mathbf{c}})\right)^2 \quad (6.7)$$

where $\tilde{\mathbf{c}}$ is a one-hot encoded digital staining matrix with the same pixel dimensions as the input image. When used in the testing phase, the one-hot encoding allows the network to generate two separate stains (H&E and the corresponding special stain) for each field of view.

The number of channels in each layer used by this deep neural network was increased by a factor of two compared to the stain transformation architecture described above to account for the larger dataset size and the need for the network to perform two distinct stain transformations.

A set of four adjacent tissue sections were used to train the virtual staining networks for H&E and the three special stains. The H&E portion of all three of the networks was trained with 1058 1424×1424-pixel images coming from 10 unique patients, the PAS network was trained with 946 1424×1424-pixel images coming from 11 unique patients, the Jones network was trained with 816 1424×1424-pixel images coming from 10 unique patients, and the Masson's Trichrome network was trained with 966 1424×1424-pixel images coming from 10 unique patients. All of the stains were validated using the same three validations slides.

Style transfer for H&E image data augmentation

In order to ensure that the stain transformation neural network is capable of being applied to a wide variety of histochemically stained H&E images, I use the CycleGAN [149] model to *augment the training dataset* by performing style transfer (**Figure 6.2**b). As discussed, these

CycleGAN networks only augment the image data used as inputs in the training phase. This CycleGAN model learns to map between two domains X and Y given the training samples x and y, where X is the domain for the original virtually stained H&E and Y is the domain for the H&E image generated by a different lab or hospital. This model performs two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. In addition, two adversarial discriminators D_X and D_Y are introduced. A diagram showing the relationship between these various networks is shown in **Figure 6.10**.

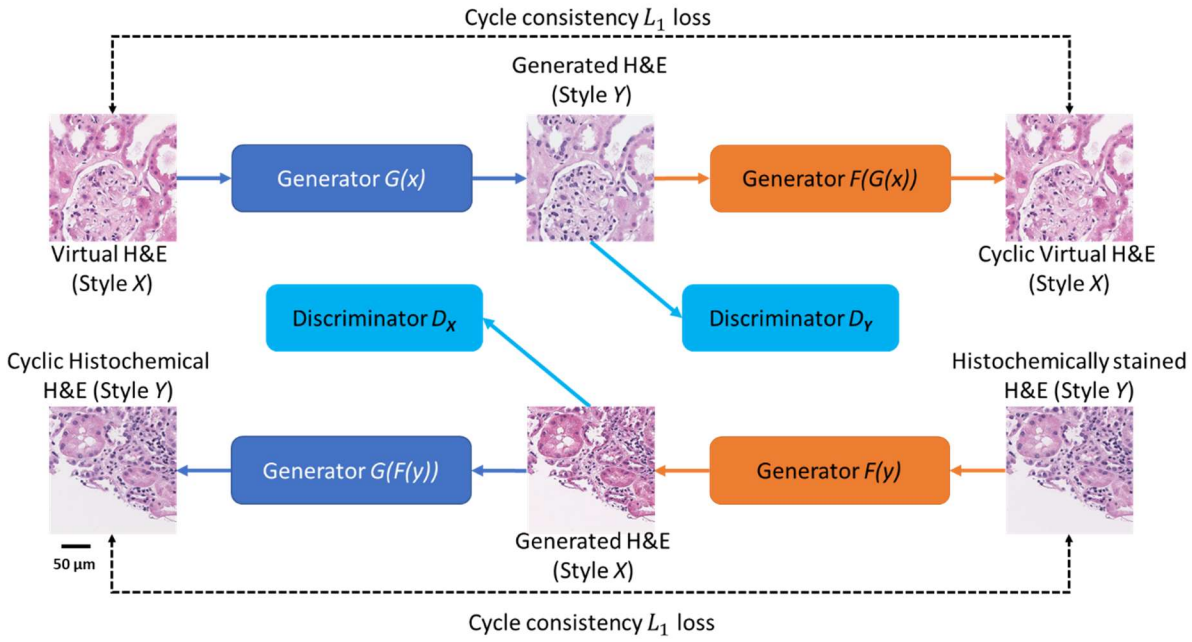

Figure 6.10 Diagram showing the transformations performed by the various networks during the training phase.

The loss function of the generator $l_{\text{generator}}$ contains two types of terms: adversarial losses l_{adv} to match the stain style of the generated images to the style of histochemically stained images in target domain; and cycle consistency losses l_{cycle} to prevent the learned mappings G and F from contradicting each other. The overall loss is therefore described by:

$$l_{\text{generator}} = \lambda \times l_{\text{cycle}} + \varphi \times l_{\text{adv}} \tag{6.8}$$

where λ and φ are relative weights/constants. For each of the networks, I set λ = 10 and φ = 1. Each generator is associated with a discriminator, which ensures that the generated image matches the distribution of the ground truth. The adversarial losses for each of the generator networks can be written as:

$$l_{\text{adv } X\to Y} = \Big(1 - D_Y\big(G(x)\big)\Big)^2 \tag{6.9}$$

$$l_{\text{adv } Y\to X} = \Big(1 - D_X\big(F(y)\big)\Big)^2 \tag{6.10}$$

And the cycle consistency loss can be described as:

$$l_{\text{cycle}} = L_1\{y, G\big(F(y)\big)\} + L_1\{x, F\big(G(x)\big)\} \tag{6.11}$$

The adversarial loss terms used to train D_X and D_Y are defined as:

$$l_{D_X} = \big(1 - D_X(x)\big)^2 + D_X\big(F(y)\big)^2 \tag{6.12}$$

$$l_{D_Y} = \big(1 - D_Y(y)\big)^2 + D_Y\big(G(x)\big)^2 \tag{6.13}$$

For these CycleGAN models, G and F use U-net architectures similar to the stain transformation network. It consists of three down-blocks followed by three up-blocks. Each of these down-blocks and up-blocks are identical to the corresponding blocks in the stain transformation network. D_X and D_Y also have similar architectures to the discriminator network of stain transformation network. However, they have four blocks rather than five blocks as in the previous model.

During the training, the Adam optimizer was used to update the learnable parameters with learning rates of 2×10^{-5} for both the generator and discriminator networks. For each step of

discriminator training, one iteration of training was performed for the generator network, and the batch size for training was set to 6.

Training of single-stain virtual staining networks

In addition to performing multiple virtual stains using a single neural network, separate networks which only generate one individual virtual stain each were also trained. These networks were used to perform the rough virtual staining that enables the elastic co-registration. These networks use the same general architecture as the stain transformation networks, with the only difference being that the first block in both the generator and the discriminator increases the number of channels to 64. The input and output images are the autofluorescence images and the histochemically stained images, respectively, processed using the image registration described in image co-registration section.

Implementation details

The image co-registration was implemented in MATLAB using version R2018a (The MathWorks Inc.). The neural networks were trained and implemented using Python version 3.6.2 with TensorFlow version 1.8.0. Timing was measured on a Windows 10 computer with two Nvidia GeForce GTX 1080 Ti GPUs, 64GB of RAM, and an Intel I9-7900X CPU.

Pathologic evaluation of kidney biopsies

An initial study of 16 sections – comparing the diagnoses made with H&E only against the diagnoses made with H&E as well as the stain-transformed special stains – was first performed to determine the feasibility of the technique. For this initial evaluation, 16 non-neoplastic kidney cases were selected by a board-certified kidney pathologist (J.E.Z.) to represent a variety of kidney diseases. For each case, the WSI of the histochemically stained H&E slide, along with a worksheet

that included a brief clinical history, were presented to 3 board-certified renal pathologists (W.D.W, M.F.P.D and A.S.). The WSIs were exported to the Zoomify format [158], and uploaded to the GIGAmacro [159] website to allow the pathologists to confidentially view the images using a standard web browser. The WSIs were viewed using standard displays (e.g., LCD Monitor, FullHD, 1920x1080 pixels).

In the diagnostic worksheet, the reviewers were given the H&E WSI and a brief patient history and asked to make a preliminary diagnosis and quantify certain features of the biopsy (i.e. number of glomeruli and arteries) and provide additional comments if necessary. After a >3-week washout period to reduce the pathologists' familiarity with the cases, the 3 reviewing pathologists received, in addition to the same histologically stained H&E WSIs and the same patient medical history, 3 computationally generated special stain WSIs for each case: MT, PAS and JMS. Being given these slides, they were asked to provide a preliminary diagnosis for a second time. This >3-week washout period was chosen to be one week greater than the College of American Pathologists Pathology and Laboratory Quality Center guidelines [160], ensuring that the pathologists were not influenced by previous diagnoses.

To test the hypothesis that using additional stain-transformed WSIs can be used to improve the preliminary diagnosis, the adjudicator pathologist (J.E.Z.) who was not among the 3 diagnosticians provided judgement to determine Concordance (C), Discordance (D) or Improvements (I) between the diagnosis quality of the first and second round of preliminary diagnoses provided by the group of diagnosticians.

To expand the total number of cases to 58 and perform the third study, (**Figure 6.3**) the same set of steps were repeated. To allow for higher throughput, in this case the WSIs were uploaded to a custom built online file viewing server based on the Orthanc server package [161]. Using this

online server, the user is able to swap between the various cases. For each case, the patient history is presented, along with the WSI and the option to swap between the various stains, where applicable. The pathologists were asked to input their diagnosis, the chronicity, and any comments that they might have into text boxes within the interface.

Once the pathologists completed the diagnoses with H&E only as well as with H&E and the stain-transformed special stains, another >3-week washout period was observed. Following this second washout period, the pathologists were given WSIs of the original histochemically stained H&E along with the 3 histochemically stained special stains coming from serial tissue sections. Two of these cases used in the preliminary study were excluded from the final analysis, as WSIs of the three special stains could not be obtained from serial tissue sections. For the first of these excluded cases, all of the pathologist's diagnoses were improved using stain-to-stain transformation, and for the second, one of the diagnoses was improved while the other two pathologists' diagnoses were concordant.

Statistical analysis

Using the preliminary study of 16 samples, a total of 41 samples were calculated to be needed in order to show statistical significance (using a power of 0.8 and an alpha level of 0.05 and using a one tailed t-test). Therefore, the total number of patients was increased 58 to ensure that the study was sufficiently powered.

A one tailed t-test was used to determine whether a statistically significant number of improvements were made when using either [H&E and stain-transformed special stains], or [H&E and histochemically stained special stains] over only [H&E] images. The statistical analyses were performed by giving a score of +1 to any improvement, -1 to any discordance and 0 to any

concordance. The score for each case was then averaged among the three pathologists who evaluated the case, and the test showed that the amount of improvement (i.e. if the average score is greater than zero) across the 58 cases was statistically significant.

A chi-squared test with two degrees of freedom was used to compare the proportion of improvements, concordances and discordances between the methods tested above. The improvements, concordances and discordances for each pathologist was compared individually.

For all tests, a P value of 0.05 or less was considered to be significant.

Chapter 7 Conclusions

Deep learning has become a key tool for image enhancement and transformations in computational microscopy. Throughout this dissertation I have demonstrated a few of the many applications of this technology.

In summary, in Chapter 1, I introduce inverse problems in microscopy. I give a short background of deep learning in relation to inverse problems, and give a general overview of how neural networks can be used to solve a number of these problems.

In Chapter 2 I show that deep learning can be used to perform super-resolution of images captured by coherent microscopes. In this chapter we demonstrate that these neural networks can both super-resolve numerical aperture-limited and pixel-limited images, showing that the technique is broadly applicable across widefield imaging systems. Chapter 3 expands upon these super-resolution results by showing that a similar framework can be applied to images captured by a scanning electron microscope. In Chapter 4, I showed an application of this technology, where it enhances and standardizes mobile phone microscope images, enabling downstream detection of sickle cell disease in thin blood smears. While in Chapter 2 through Chapter 4 I showed a few different modalities that can be enhanced using deep learning, these general techniques can be generalized beyond to fluorescence microscopy [41] and beyond [12].

In Chapter 5 I introduce a neural network that can be used to generate multiple stains from a single tissue section. I show that this technology generates highly accurate stains, while enabling new features such as micro structured staining within a single field of view, as well as stain blending. Chapter 6 demonstrates how a network can be trained to perform transformations from H&E to various special stains by using the multistain network proposed in the previous chapter to

generate pairs of synthetic data. This technology was then shown to be effective at improving preliminary diagnoses as well as statistically equivalent in quality to histochemical staining of special stains.

Looking forward, there is still significant room for these technologies to continue to develop. The field of deep learning is rapidly growing, with continual improvements being made to training techniques and network architectures. Each of the applications presented here can be improved upon through both hardware and software advancements. Furthermore, all of the networks presented here are simply a proof of concept. Larger datasets will likely improve the quality of the networks, and significantly more robust models can be used to ensure that they generalize beyond individual microscopes or samples. The techniques may also be improved by improvements in the dataset creation workflow such as improved registration algorithms.

References

1. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. (Curran Associates, Inc., 2012), pp. 1097–1105.
2. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *Proceedings of the International Conference on Learning Representations (ICLR)* (2015).
3. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.
4. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (Springer International Publishing, 2015), pp. 234–241.
5. W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bulletin of Mathematical Biophysics **5**, 115–133 (1943).
6. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE **86**, 2278–2324 (1998).
7. C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence **38**, 295–307 (2016).
8. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437–1443 (2017).

9. Y. Rivenson, Y. Zhang, H. Gunaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Science & Applications **7**, e17141 (n.d.).
10. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-To-Image Translation With Conditional Adversarial Networks," in (2017), pp. 1125–1134.
11. I. J. 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 (NIPS)*, NIPS'14 (MIT Press, 2014).
12. K. de Haan, Y. Rivenson, Y. Wu, and A. Ozcan, "Deep-Learning-Based Image Reconstruction and Enhancement in Optical Microscopy," Proceedings of the IEEE **108**, 30–50 (2020).
13. T. Liu, K. de Haan, Y. Rivenson, Z. Wei, X. Zeng, Y. Zhang, and A. Ozcan, "Deep learning-based super-resolution in coherent imaging systems," Scientific Reports **9**, 3926 (2019).
14. G. Popescu, *Quantitative Phase Imaging of Cells and Tissues* (McGraw Hill Professional, 2011).
15. D. Gabor, "A New Microscopic Principle," Nature **161**, 777–778 (1948).
16. E. N. Leith and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory*," J. Opt. Soc. Am., JOSA **52**, 1123–1130 (1962).
17. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
18. 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).

19. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," Journal of the Optical Society of America **73**, 1434 (1983).

20. L. J. Allen and M. P. Oxley, "Phase retrieval from series of images obtained by defocus variation," Optics Communications **199**, 65–75 (2001).

21. I. Yamaguchi and T. Zhang, "Phase-shifting digital holography," Opt. Lett., OL **22**, 1268–1270 (1997).

22. Z. Wang, L. Millet, M. Mir, H. Ding, S. Unarunotai, J. Rogers, M. U. Gillette, and G. Popescu, "Spatial light interference microscopy (SLIM)," Opt Express **19**, 1016–1026 (2011).

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

24. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," (2017).

25. Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. De Haan, and A. Ozcan, "Cross-modality deep learning brings bright-field microscopy contrast to holography," arXiv:1811.07103 [physics] (2018).

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

27. V. Mico, Z. Zalevsky, P. García-Martínez, and J. García, "Superresolved imaging in digital holography by superposition of tilted wavefronts," Appl. Opt., AO **45**, 822–828 (2006).

28. V. Mico, Z. Zalevsky, P. García-Martínez, and J. García, "Synthetic aperture superresolution with multiple off-axis holograms," J. Opt. Soc. Am. A, JOSAA **23**, 3162–3170 (2006).

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

30. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nature Photonics **7**, 739–745 (2013).

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

32. A. Meiri, E. Gur, J. Garcia, V. Micó, B. Javidi, and Z. Zalevsky, "Super Resolved Holographic Configurations," Multi-Dimensional Imaging 225–239 (2014).

33. M. Paturzo, F. Merola, S. Grilli, S. De Nicola, A. Finizio, and P. Ferraro, "Super-resolution in digital holography by a two-dimensional dynamic phase grating," Optics express **16**, 17107–17118 (2008).

34. A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, "Sparsity-based single-shot subwavelength coherent diffractive imaging," Nat Mater **11**, 455–459 (2012).

35. Y. Liu, L. Tian, C.-H. Hsieh, and G. Barbastathis, "Compressive holographic two-dimensional localization with $1/30^2$ subpixel accuracy," Opt. Express, OE **22**, 9774–9782 (2014).

36. Y. Rivenson, M. A. Shalev, and Z. Zalevsky, "Compressive Fresnel holography approach for high-resolution viewpoint inference," Optics Letters **40**, 5606 (2015).

37. C. Fournier, F. Jolivet, L. Denis, N. Verrier, E. Thiebaut, C. Allier, and T. Fournel, "Pixel super-resolution in digital holography by regularized reconstruction," Applied optics **56**, 69–77 (2017).

38. 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).

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

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

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

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

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

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

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

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

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

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

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

50. 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*, Y. Bengio and Y. LeCun, eds. (2015).

51. K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, and A. Ozcan, "Resolution enhancement in scanning electron microscopy using deep learning," Scientific Reports **9**, 12050 (2019).

52. A. E. Vladár, M. T. Postek, and B. Ming, "On the Sub-Nanometer Resolution of Scanning Electron and Helium Ion Microscopes," Microscopy Today **17**, 6–13 (2009).

53. J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. M. Ritchie, J. H. J. Scott, and D. C. Joy, *Scanning Electron Microscopy and X-Ray Microanalysis* (Springer, 2017).

54. H. Schatten, "Low voltage high-resolution SEM (LVHRSEM) for biological structural and molecular analysis," Micron **42**, 175–185 (2011).

55. W. R. Pelton, "Distinguishing the Cause of Textile Fiber Damage Using the Scanning Electron Microscope (SEM)," Journal of Forensic Sciences **40**, 874–882 (1995).

56. H. Wu and S. N. Melkote, "Study of Ductile-to-Brittle Transition in Single Grit Diamond Scribing of Silicon: Application to Wire Sawing of Silicon Wafers," J. Eng. Mater. Technol **134**, 041011 (2012).

57. Z. Zhang, B. Wang, R. Kang, B. Zhang, and D. Guo, "Changes in surface layer of silicon wafers from diamond scratching," CIRP Annals **64**, 349–352 (2015).

58. Z. Zhang, D. Guo, B. Wang, R. Kang, and B. Zhang, "A novel approach of high speed scratching on silicon wafers at nanoscale depths of cut," Scientific Reports **5**, 16395 (2015).

59. B. Wang, Z. Zhang, K. Chang, J. Cui, A. Rosenkranz, J. Yu, C.-T. Lin, G. Chen, K. Zang, J. Luo, N. Jiang, and D. Guo, "New Deformation-Induced Nanostructure in Silicon," Nano Lett. **18**, 4611–4617 (2018).

60. B. Titze, "Techniques to prevent sample surface charging and reduce beam damage effects for SBEM imaging," Dissertation (2013).

61. C. G. Golding, L. L. Lamboo, D. R. Beniac, and T. F. Booth, "The scanning electron microscope in microbiology and diagnosis of infectious disease," Scientific Reports **6**, (2016).

62. R. F. Egerton, P. Li, and M. Malac, "Radiation damage in the TEM and SEM," Micron **35**, 399–409 (2004).

63. S. Sreehari, S. V. Venkatakrishnan, K. L. Bouman, J. P. Simmons, L. F. Drummy, and C. A. Bouman, "Multi-resolution Data Fusion for Super-Resolution Electron Microscopy," in *The IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2017).

64. S. Tsiper, O. Dicker, I. Kaizerman, Z. Zohar, M. Segev, and Y. C. Eldar, "Sparsity-Based Super Resolution for SEM Images," Nano Lett. **17**, 5437–5445 (2017).

65. F. Yano and S. Nomura, "Deconvolution of scanning electron microscopy images," Scanning **15**, 19–24 (1993).

66. J. Roels, J. Aelterman, J. D. Vylder, H. Luong, Y. Saeys, and W. Philips, "Bayesian deconvolution of scanning electron microscopy images using point-spread function estimation and non-local regularization," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2016).

67. Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, "Electron ptychography of 2D materials to deep sub-ångström resolution," Nature **559**, 343 (2018).

68. M. J. Humphry, B. Kraus, A. C. Hurst, A. M. Maiden, and J. M. Rodenburg, "Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging," Nature Communications **3**, 730 (2012).

69. 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," Nature Methods **15**, 263–266 (2018).

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

71. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D: Nonlinear Phenomena **60**, 259–268 (1992).

72. K. de Haan, H. Ceylan Koydemir, Y. Rivenson, D. Tseng, E. Van Dyne, L. Bakic, D. Karinca, K. Liang, M. Ilango, E. Gumustekin, and A. Ozcan, "Automated screening of sickle cells using a smartphone-based microscope and deep learning," npj Digital Medicine **3**, 1–9 (2020).

73. CDC, "Data & Statistics on Sickle Cell Disease | CDC," https://www.cdc.gov/ncbddd/sicklecell/data.html.

74. B. Modell and M. Darlison, "Global epidemiology of haemoglobin disorders and derived service indicators," Bull. World Health Organ. **86**, 480–487 (2008).

75. "Sickle Cell Disease," https://www.afro.who.int/health-topics/sickle-cell-disease.

76. A. H. Al-Salem, "Splenic Complications of Sickle Cell Anemia and the Role of Splenectomy," ISRN Hematol **2011**, 864257 (2011).

77. X. Yang, J. Kanter, N. Z. Piety, M. S. Benton, S. M. Vignes, and S. S. Shevkoplyas, "A simple, rapid, low-cost diagnostic test for sickle cell disease," Lab Chip **13**, 1464–1467 (2013).

78. T. Adeyemo, O. Ojewunmi, and A. Oyetunji, "Evaluation of high performance liquid chromatography (HPLC) pattern and prevalence of beta-thalassaemia trait among sickle cell disease patients in Lagos, Nigeria," Pan Afr Med J **18**, 71–71 (2014).

79. H. Wajcman and K. Moradkhani, "Abnormal haemoglobins: detection & characterization," Indian J Med Res **134**, 538–546 (2011).

80. J. K. Fleming, "Evaluation of HbA1c on the Roche COBAS Integra 800 closed tube system," Clinical Biochemistry **40**, 822–827 (2007).

81. N. Z. Piety, A. George, S. Serrano, M. R. Lanzi, P. R. Patel, M. P. Noli, S. Kahan, D. Nirenberg, J. F. Camanda, G. Airewele, and S. S. Shevkoplyas, "A Paper-Based Test for Screening Newborns for Sickle Cell Disease," Scientific Reports **7**, 45488 (2017).

82. S. M. Knowlton, I. Sencan, Y. Aytar, J. Khoory, M. M. Heeney, I. C. Ghiran, and S. Tasoglu, "Sickle cell detection using a smartphone," Scientific Reports **5**, 15022 (2015).

83. Y. Alapan, A. Fraiwan, E. Kucukal, M. N. Hasan, R. Ung, M. Kim, I. Odame, J. A. Little, and U. A. Gurkan, "Emerging Point-of-Care Technologies for Sickle Cell Disease Screening and Monitoring," Expert Rev Med Devices **13**, 1073–1093 (2016).

84. M. Bond, B. Hunt, B. Flynn, P. Huhtinen, R. Ware, and R. Richards-Kortum, "Towards a point-of-care strip test to diagnose sickle cell anemia," PLOS ONE **12**, e0177732 (2017).

85. C. Steele, A. Sinski, J. Asibey, M.-D. Hardy-Dessources, G. Elana, C. Brennan, I. Odame, C. Hoppe, M. Geisberg, E. Serrao, and C. T. Quinn, "Point-of-care screening for sickle cell disease in low-resource settings: A multi-center evaluation of HemoTypeSC, a novel rapid test," Am. J. Hematol. **94**, 39–45 (2019).

86. O. Nnodu, H. Isa, M. Nwegbu, C. Ohiaeri, S. Adegoke, R. Chianumba, N. Ugwu, B. Brown, J. Olaniyi, E. Okocha, J. Lawson, A.-A. Hassan, I. Diaku-Akinwumi, A. Madu, O. Ezenwosu, Y. Tanko, U. Kangiwa, A. Girei, Y. Israel-Aina, A. Ladu, P. Egbuzu, U. Abjah, A. Okolo, N. Akbulut-Jeradi, M. Fernandez, F. B. Piel, and A. Adekile, "HemoTypeSC, a low-cost point-of-care testing device for sickle cell disease: Promises and challenges," Blood Cells, Molecules, and Diseases **78**, 22–28 (2019).

87. M. Poostchi, K. Silamut, R. J. Maude, S. Jaeger, and G. Thoma, "Image analysis and machine learning for detecting malaria," Translational Research **194**, 36–55 (2018).

88. J. Ford, "Red blood cell morphology," Int J Lab Hematol **35**, 351–357 (2013).

89. A. Adewoyin and B. Nwogoh, "PERIPHERAL BLOOD FILM - A REVIEW," Ann Ib Postgrad Med **12**, 71–79 (2014).

90. D. C. Rees, T. N. Williams, and M. T. Gladwin, "Sickle-cell disease," The Lancet **376**, 2018–2031 (2010).

91. M. Xu, D. P. Papageorgiou, S. Z. Abidi, M. Dao, H. Zhao, and G. E. Karniadakis, "A deep convolutional neural network for classification of red blood cells in sickle cell anemia," PLOS Computational Biology **13**, e1005746 (2017).

92. M. Zhang, X. Li, M. Xu, and Q. Li, "RBC Semantic Segmentation for Sickle Cell Disease Based on Deformable U-Net," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (Springer International Publishing, 2018), pp. 695–702.

93. H. Zhu, S. O. Isikman, O. Mudanyali, A. Greenbaum, and A. Ozcan, "Optical imaging techniques for point-of-care diagnostics," Lab Chip **13**, 51–67 (2013).

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

95. S. K. Vashist, P. B. Luppa, L. Y. Yeo, A. Ozcan, and J. H. T. Luong, "Emerging Technologies for Next-Generation Point-of-Care Testing," Trends Biotechnol. **33**, 692–705 (2015).

96. 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).

97. H. Zhu, I. Sencan, J. Wong, S. Dimitrov, D. Tseng, K. Nagashima, and A. Ozcan, "Cost-effective and rapid blood analysis on a cell-phone," Lab Chip **13**, 1282–1288 (2013).

98. Q. Wei, H. Qi, W. Luo, D. Tseng, S. J. Ki, Z. Wan, Z. Göröcs, L. A. Bentolila, T.-T. Wu, R. Sun, and A. Ozcan, "Fluorescent imaging of single nanoparticles and viruses on a smart phone," ACS Nano **7**, 9147–9155 (2013).

99. 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**, 1–8 (2017).

100. 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).

101. T. Laksanasopin, T. W. Guo, S. Nayak, A. A. Sridhara, S. Xie, O. O. Olowookere, P. Cadinu, F. Meng, N. H. Chee, J. Kim, C. D. Chin, E. Munyazesa, P. Mugwaneza, A. J. Rai, V. Mugisha, A. R. Castro, D. Steinmiller, V. Linder, J. E. Justman, S. Nsanzimana, and S. K. Sia, "A smartphone dongle for diagnosis of infectious diseases at the point of care," Science Translational Medicine **7**, 273re1-273re1 (2015).

102. T. Guo, R. Patnaik, K. Kuhlmann, A. J. Rai, and S. K. Sia, "Smartphone dongle for simultaneous measurement of hemoglobin concentration and detection of HIV antibodies," Lab Chip **15**, 3514–3520 (2015).

103. S. Lee, D. O'Dell, J. Hohenstein, S. Colt, S. Mehta, and D. Erickson, "NutriPhone: a mobile platform for low-cost point-of-care quantification of vitamin B_{12} concentrations," Scientific Reports **6**, 28237 (2016).

104. 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," Scientific Reports **6**, 39203 (2016).

105. T. S. Park, W. Li, K. E. McCracken, and J.-Y. Yoon, "Smartphone quantifies Salmonella from paper microfluidics," Lab Chip **13**, 4832–4840 (2013).

106. R. Lin, A. Skandarajah, R. E. Gerver, H. D. Neira, D. A. Fletcher, and A. E. Herr, "A lateral electrophoretic flow diagnostic assay," Lab Chip **15**, 1488–1496 (2015).

107. M. V. D'Ambrosio, M. Bakalar, S. Bennuru, C. Reber, A. Skandarajah, L. Nilsson, N. Switz, J. Kamgno, S. Pion, M. Boussinesq, T. B. Nutman, and D. A. Fletcher, "Point-of-care

quantification of blood-borne filarial parasites with a mobile phone microscope," Science Translational Medicine **7**, 286re4-286re4 (2015).

108. G. T. Smith, L. Li, Y. Zhu, and A. K. Bowden, "Low-power, low-cost urinalysis system with integrated dipstick evaluation and microscopic analysis," Lab Chip **18**, 2111–2123 (2018).
109. 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).
110. O. Holmström, N. Linder, B. Ngasala, A. Mårtensson, E. Linder, M. Lundin, H. Moilanen, A. Suutala, V. Diwan, and J. Lundin, "Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths and Schistosoma haematobium," Global Health Action **10**, 1337325 (2017).
111. Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing **13**, 600–612 (2004).
112. S. Mavandadi, S. Feng, F. Yu, S. Dimitrov, K. Nielsen-Saines, W. R. Prescott, and A. Ozcan, "A mathematical framework for combining decisions of multiple experts toward accurate and remote diagnosis of malaria using tele-microscopy," PLoS One **7**, e46192–e46192 (2012).
113. A. Ozcan, "Educational Games for Malaria Diagnosis," Science Translational Medicine **6**, 233ed9-233ed9 (2014).
114. O. Alvarez, N. S. Montague, M. Marin, R. O'Brien, and M. M. Rodriguez, "Quantification of sickle cells in the peripheral smear as a marker of disease severity," Fetal Pediatr Pathol **34**, 149–154 (2015).

115. W. H. Organization and S. WHO Expert Committee on the Control and Surveillance of Human African Trypanosomiasis (2013: Geneva, *Control and Surveillance of Human African Trypanosomiasis: Report of a WHO Expert Committee* (World Health Organization, 2013).

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

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

118. 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: Science & Applications **9**, 78 (2020).

119. M. Titford, "The long history of hematoxylin," Biotechnic & Histochemistry **80**, 73–78 (2005).

120. H. A. Alturkistani, F. M. Tashkandi, and Z. M. Mohammedsaleh, "Histological Stains: A Literature Review and Case Study," Glob J Health Sci **8**, 72–79 (2016).

121. S. Yamabayashi, "Periodic acid-Schiff-alcian blue: a method for the differential staining of glycoproteins," Histochem. J. **19**, 565–571 (1987).

122. 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," PNAS **111**, 15304–15309 (2014).

123. 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**, (2017).

124. 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," Nature Photonics **10**, 534–540 (2016).

125. S. González and Y. Gilaberte-Calzada, "In vivo reflectance-mode confocal microscopy in clinical dermatology and cosmetology," International Journal of Cosmetic Science **30**, 1–17 (2008).

126. D. Mayerich, M. J. Walsh, A. Kadjacsy-Balla, P. S. Ray, S. M. Hewitt, and R. Bhargava, "Stain-less staining for computed histopathology," Technology (Singap World Sci) **3**, 27–31 (2015).

127. T. Yoshitake, M. G. Giacomelli, L. M. Quintana, H. Vardeh, L. C. Cahill, B. E. Faulkner-Jones, J. L. Connolly, D. Do, and J. G. Fujimoto, "Rapid histopathological imaging of skin and breast cancer surgical specimens using immersion microscopy with ultraviolet surface excitation," Sci Rep **8**, 1–12 (2018).

128. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl,

L. L. Rubin, P. Nelson, and S. Finkbeiner, "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images," Cell **173**, 792–803 (2018).

129. Y. Wu, Y. Luo, G. Chaudhari, Y. Rivenson, A. Calis, K. Haan, and A. Ozcan, "Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram," Light: Science & Applications **8**, 25 (2019).
130. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," Light: Science & Applications **8**, 23 (2019).
131. N. Borhani, A. J. Bower, S. A. Boppart, and D. Psaltis, "Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy," Biomed. Opt. Express, BOE **10**, 1339–1350 (2019).
132. W. Bulten and G. Litjens, "Unsupervised Prostate Cancer Detection on H&E using Convolutional Adversarial Autoencoders," arXiv:1804.07098 [cs] (2018).
133. A. Bentaieb and G. Hamarneh, "Adversarial Stain Transfer for Histopathology Image Analysis," IEEE Transactions on Medical Imaging **37**, 792–802 (2018).
134. A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and Eosin Staining of Tissue and Cell Sections," Cold Spring Harb Protoc **2008**, pdb.prot4986 (2008).
135. G. A. Herrera and R. L. Lott, "Silver Stains In Diagnostic Renal Pathology," Journal of Histotechnology **19**, 219–223 (1996).
136. K. Wada, *Labelme: Image Polygonal Annotation with Python* (2016).
137. K. de Haan, Y. Zhang, J. E. Zuckerman, T. Liu, A. E. Sisk, M. F. P. Diaz, K.-Y. Jen, A. Nobori, S. Liou, S. Zhang, R. Riahi, Y. Rivenson, W. D. Wallace, and A. Ozcan, "Deep

learning-based transformation of H&E stained tissues into special stains," Nat Commun **12**, 4884 (2021).

138. *Global Tissue Diagnostics Market, Forecast to 2022* (2018).

139. P. D. Walker, T. Cavallo, S. M. Bonsib, and Ad Hoc Committee on Renal Biopsy Guidelines of the Renal Pathology Society, "Practice guidelines for the renal biopsy," Mod. Pathol. **17**, 1555–1563 (2004).

140. 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," Nature Biomedical Engineering **1**, 957–966 (2017).

141. 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," Nature Biomedical Engineering **1**, 1–10 (2017).

142. N. Bayramoglu, M. Kaakinen, L. Eklund, and J. Heikkilä, "Towards Virtual H and E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks," in *IEEE International Conference on Computer Vision Workshops (ICCVW)* (2017), pp. 64–71.

143. A. Rana, A. Lowe, M. Lithgow, K. Horback, T. Janovitz, A. D. Silva, H. Tsai, V. Shanmugam, A. Bayat, and P. Shah, "Use of Deep Learning to Develop and Analyze Computational Hematoxylin and Eosin Staining of Prostate Core Biopsy Images for Tumor Diagnosis," JAMA Netw Open **3**, e205111–e205111 (2020).

144. S. Roy-Chowdhuri, S. Dacic, M. Ghofrani, P. B. Illei, L. J. Layfield, C. Lee, C. W. Michael, R. A. Miller, J. W. Mitchell, B. Nikolic, J. A. Nowak, N. J. Pastis, C. A. Rauch, A. Sharma,

L. Souter, B. L. Billman, N. E. Thomas, P. A. VanderLaan, J. S. Voss, M. M. Wahidi, L. B. Yarmus, and C. R. Gilbert, "Collection and Handling of Thoracic Small Biopsy and Cytology Specimens for Ancillary Studies: Guideline From the College of American Pathologists in Collaboration With the American College of Chest Physicians, Association for Molecular Pathology, American Society of Cytopathology, American Thoracic Society, Pulmonary Pathology Society, Papanicolaou Society of Cytopathology, Society of Interventional Radiology, and Society of Thoracic Radiology," Arch. Pathol. Lab. Med. (2020).

145. "Preliminary Evaluation of the Utility of Deep Generative Histopathology Image Translation at a Mid-Sized NCI Cancer Center | bioRxiv," https://www.biorxiv.org/content/10.1101/2020.01.07.897801v2.

146. A. Lahiani, I. Klaman, N. Navab, S. Albarqouni, and E. Klaiman, "Seamless Virtual Whole Slide Image Synthesis and Validation Using Perceptual Embedding Consistency," IEEE J Biomed Health Inform (2020).

147. M. Gadermayr, V. Appel, B. M. Klinkhammer, P. Boor, and D. Merhof, "Which Way Round? A Study on the Performance of Stain-Translation for Segmenting Arbitrarily Dyed Histological Images," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds., Lecture Notes in Computer Science (Springer International Publishing, 2018), pp. 165–173.

148. A. Kapil, T. Wiestler, S. Lanzmich, A. Silva, K. Steele, M. Rebelatto, G. Schmidt, and N. Brieu, "DASGAN -- Joint Domain Adaptation and Segmentation for the Analysis of Epithelial Regions in Histopathology PD-L1 Images," arXiv:1906.11118 [cs, eess] (2019).

149. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2242–2251.

150. J. P. Cohen, M. Luck, and S. Honari, "Distribution Matching Losses Can Hallucinate Features in Medical Image Translation," arXiv:1805.08841 [cs] (2018).

151. M. Fujitani, Y. Mochizuki, S. Iizuka, E. Simo-Serra, H. Kobayashi, C. Iwamoto, K. Ohuchida, M. Hashizume, H. Hontani, and H. Ishikawa, "Re-staining Pathology Images by FCNN," in *2019 16th International Conference on Machine Vision Applications (MVA)* (2019), pp. 1–6.

152. C. Mercan, G. Reijnen-Mooij, D. T. Martin, J. Lotz, N. Weiss, M. van Gerven, and F. Ciompi, "Virtual staining for mitosis detection in Breast Histopathology," arXiv:2003.07801 [cs, eess] (2020).

153. T. W. Bauer, L. Schoenfield, R. J. Slaw, L. Yerian, Z. Sun, and W. H. Henricks, "Validation of whole slide imaging for primary diagnosis in surgical pathology," Arch. Pathol. Lab. Med. **137**, 518–524 (2013).

154. M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, "Staingan: Stain Style Transfer for Digital Histological Images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (IEEE, 2019), pp. 953–956.

155. V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," Scientific Reports **9**, 16884 (2019).

156. R. A. ERLANDSON, "Role of Electron Microscopy in Modern Diagnostic Surgical Pathology," Modern Surgical Pathology 71–84 (2009).

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

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

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

160. L. Pantanowitz, J. H. Sinard, W. H. Henricks, L. A. Fatheree, A. B. Carter, L. Contis, B. A. Beckwith, A. J. Evans, A. Lal, A. V. Parwani, and College of American Pathologists Pathology and Laboratory Quality Center, "Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center," Arch Pathol Lab Med **137**, 1710–1722 (2013).

161. S. Jodogne, "The Orthanc Ecosystem for Medical Imaging," J Digit Imaging **31**, 341–352 (2018).