STRUCTURED LIGHT FOR DATA-DRIVEN QUANTITATIVE TISSUE IMAGING

by
Mason Tianyi Chen

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March, 2022

Abstract

Compared to the other major medical imaging modalities (CT, MRI, PET, ultrasound, and microscopy), conventional wide-field optical imaging is particularly qualitative. For endoscopic and laparoscopic optical imaging, the spatial scale of the image is typically unknown, heterogeneous, and rapidly changing frame-to-frame. Moreover, the fundamental optical properties of the imaged tissue are conflated in the acquired images. This leads to several clinical limitations. The absorption coefficient, which can indicate important biomarkers such as tissue oxygenation, cannot be unambiguously determined. Moreover, for machine learning analysis of endoscopic images, qualitative image inputs lead to large training datasets requirements and limit generalizability compared to algorithms trained with quantitative images. Quantitative optical imaging may be key to enabling reliable artificial intelligence algorithms for tasks such as tissue classification or tumor margin assessment.

Spatial Frequency Domain Imaging (SFDI) is a relatively new optical imaging technique that is capable of wide-field quantification and mapping of tissue optical properties. Using structured illumination at different phases and spatial frequencies, it is capable of unambiguously decoupling optical absorption and scattering. In recent

years, SFDI has seen growing use in many applications, such as wound monitoring and diabetic ulcer staging. Unfortunately, conventional SFDI approaches suffer from the trade-off between imaging speed and accuracy. SFDI also requires a projection system, which limits the potential for clinical adoption in space-constrained applications, such as endoscopic imaging.

In this work, we present a technique that leverages the power of data-driven methods, such as convolutional neural networks, to achieve real-time and accurate optical property and tissue oxygenation mapping from single-shot SFDI images. We also develop a signal processing model for projector-free SFDI using random laser speckle patterns as structured illumination. This approach is promising for rapid, low-cost, and compact endoscopic imaging of optical biomarkers. Overall, our work has the potential to facilitate clinical translation and adoption of quantitative tissue imaging techniques.

Thesis Committee Members

Dr. Nicholas Durr (Primary Advisor)
Assistant Professor
Department of Biomedical Engineering
Johns Hopkins University

Dr. Jin Kang
Professor
Department of Electrical and Computer Engineering
Johns Hopkins University

Dr. Ji Yi
Assistant Professor

Department of Biomedical Engineering
Johns Hopkins University

Acknowledgements

First and foremost, I would like to thank my advisor Dr. Nicholas Durr. The past five years would not have been possible without his mentorship and inspiration. I will always remember our first meeting when he told me that he did not want students to work *for* him but to work *with* him. As a mentor, Nick respects me as a scientist and gives me the freedom to choose my own research direction and career path. I am extremely grateful to have had his constant guidance and support on every aspect of my graduate career. I would also like to thank Dr. Jin Kang and Dr. Ji Yi for serving on my committee and providing valuable feedback.

I want to thank our lab alumni Faisal and Jordan who helped me tremendously in the first two years of my PhD, and everyone in our lab who have supported me along the way - Greg, Taylor, Ryan, Mayank, Luojie, Marisa, and Anisha. I especially want to thank my lab mates and friends Greg and Taylor. As the first group of students in Nick's lab, we have learned and grown so much together, and this journey would not have been the same without their kind advice and encouragement.

I must also thank my mother who has given me unconditional love and support, and my family in the U.S. for always making me feel at home. I hope I have made

you proud. Finally, I would like to say thank you to the friends that I have made over my years at Hopkins, and the old friends from high school and college. From first coming to the States in 2013 to now earning a PhD degree in Biomedical Engineering, the past nine years were not without ups and downs. I am grateful for the friends and family who made this a memory that I will cherish forever.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Technical Background and Significance

1.1.1 Optical Properties of Biological Tissues

Light-tissue interactions in biological tissues primarily involve the absorption and scattering of photons [8] (Fig. 1-1). Absorption is due to the presence of chromophores, which are molecules that absorb particular wavelengths of light, such as hemoglobin, water, fat, and melanin. Scattering is caused by spatial discontinuities in the index of refraction that arise from tissue microstructures, including collagen, nuclei, and mitochondria.

The optical properties of biological tissues, including absorption (μ_a) and scattering coefficients (μ_s), define how tissues interact with photons. The units of μ_a and μ_s are inverse length, such as mm^{-1} or cm^{-1}. The absorption coefficient is related to the

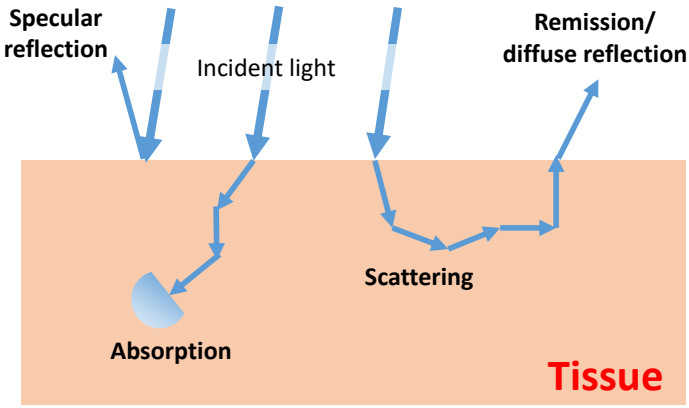

Figure 1-1. The primary interactions between biological tissues and light are photon absorption and scattering. Light can be specularly reflected at the tissue interface or diffusely reflected and remitted through scattering.

density of photon-absorbing chromophores (ρ_a) and the absorption cross-section (σ_a):

$$\mu_a = \rho_a \sigma_a. \tag{1.1}$$

σ_a is defined as:

$$\sigma_a = Q_a A_a, \tag{1.2}$$

where Q_a is the absorption efficiency and A_a is the geometrical cross-section of chromophores. Similarly, μ_s is defined as:

$$\mu_s = \rho_s \sigma_s. \tag{1.3}$$

ρ_s and σ_s denote the volume density and effective cross-section of scattering

particles, respectively, and the latter can be expressed as:

$$\sigma_s = Q_s A_s. \tag{1.4}$$

Q_s is the scattering efficiency and A_s is the geometrical size of scatterers. Note that both Q_a and Q_s are dimensionless, and σ can be smaller or larger than A [9].

Another useful optical property is the reduced scattering coefficient (μ_s'):

$$\mu_s' = \mu_s(1-g), \tag{1.5}$$

where g is the anisotropy. The concept of anisotropy is described in Fig. 1-2. Photons move in multiple small steps of size $1/\mu_s$, or mean free paths (mfp), at partial deflection angles θ before being absorbed or remitted by tissues, and anisotropy is defined as the average cosine of deflection angles:

$$g = < cos\theta > . \tag{1.6}$$

μ_s' is a lumped property that describes the diffusion of photons in a random walk of step size of $1/\mu_s'$ (reduced mean free path, or mfp') at isotropic scattering [9]. Because we assume that tissues are semi-infinite media and light behaves as a ballistic photon, μ_a and μ_s (or μ_s') are only an approximation to the true light-tissue interactions and do not model other phenomena, such as transmission, diffraction, and interference.

As previously discussed, photon absorption is due to the presence of chromophores,

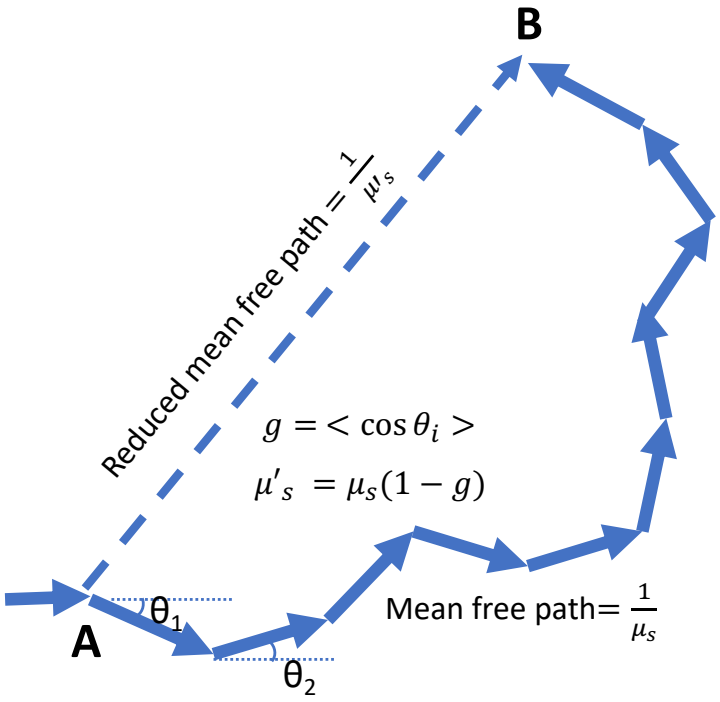

Figure 1-2. Photons take many "mean free paths" (mfp) at partial deflection angles before being absorbed or remitted. Anisotropy (g) is defined as the average cosine of the partial deflection angles. The reduced scattering coefficient is a lumped property incorporating the scattering coefficient and anisotropy, which models photon movement as one "reduced mean free path" (mfp').

the absorbance of which depends on the wavelength of light (Fig. 1-3) [1]. Concentrations of various chromophores, such as water and oxygenated and de-oxygenated hemoglobin, are clinically useful biomarkers for patient monitoring and tissue type delineation [10, 11]. According to the Beer-Lambert law, the absorption coefficient is related to chromophore concentrations:

$$\mu_a(\lambda_i) = \sum_{n=1}^{N} \epsilon_n(\lambda_i) c_n. \tag{1.7}$$

$\epsilon_n(\lambda_i)$ stands for the extinction coefficient of chromophore n at wavelength λ_i, c_n is

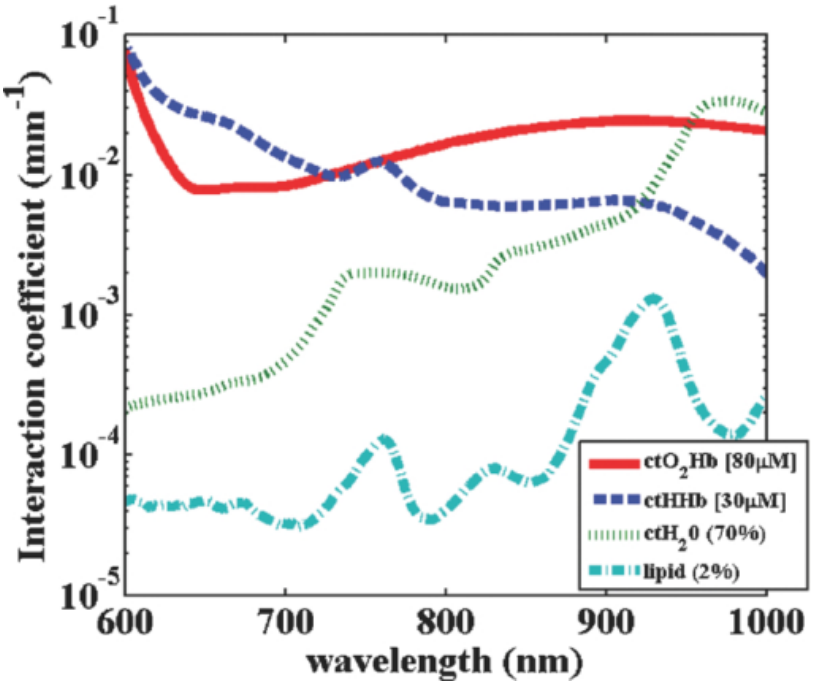

Figure 1-3. Extinction coefficients of common chromophores in the near-infrared region. Figure adapted from [1].

its concentration, and N the total number of chromophores. If absorption coefficients at different wavelengths are known, one can calculate the chromophore concentrations by solving a system of linear equations.

1.1.2 Importance of Optical Properties as Tissue Biomarkers

The optical properties of tissues affect both diagnostic and therapeutic applications of light [12]. For example, light can penetrate and probe tissues at varying depths depending on the optical properties. Quantifying the tissue optical properties by analyzing the remitted signals is critical for diagnostic applications. The deposition of energy using the absorption coefficients is central to many therapeutic applications of light. Therefore, the modeling and estimation of the optical properties of biological

tissues have been an ongoing area of research for many decades.

In this dissertation, we are particular interested in quantitative tissue characterization using optical properties. Optical properties can be useful clinical biomarkers for measuring trends and detecting abnormalities in tissue metabolism, tissue oxygenation, and cellular proliferation [13–17]. More specifically, one can use the absorption coefficients at multiple wavelengths to measure the tissue oxygenation or perfusion, which is integral for surgical guidance and patient monitoring [18, 7]. The scattering coefficients can also be used to enhance contrast between tissue types in functional or structural imaging [19, 20, 10].

Another important aspect of optical properties is that they are a quantitative metric that describes light-tissue interactions independent of external environment. Conventional wide-field images are qualitative, and the image color and intensity depend on a range of factors. For example, images of the same slab of tissue could look different due to the relative positions between light source and camera and the imaging hardware. The appearance of the image also depends on surface parameters of the sample and the underlying optical properties. However, only the optical properties characterize the intrinsic physical properties of the tissue. In fact, μ_a and μ_s' are absolute measurements that can be directly compared across imaging platforms, study sites, and time scales. Thus, quantitative imaging of tissue optical properties can facilitate more objective, precise, and optimized management of patients [3].

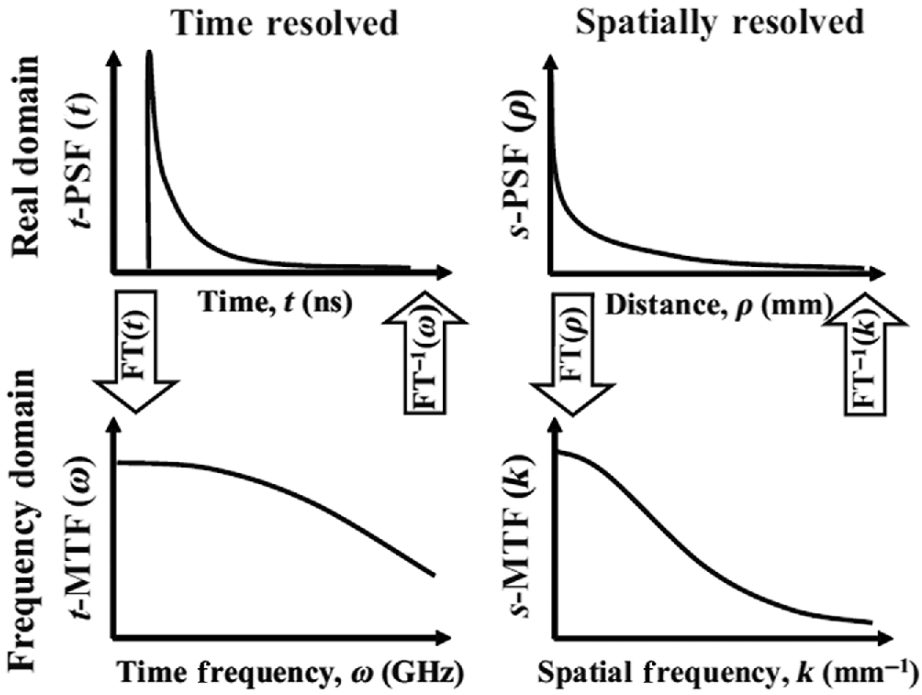

Figure 1-4. Four domains of optical property measurement techniques: real temporal domain, real spatial domain, temporal frequency domain, and spatial frequency domain. Real and temporal domains are a Fourier transform pair. Figure adapted from [2].

1.2 Optical Property Measurement Techniques

1.2.1 Four Measurement Domains

Separating optical absorption from scattering is a challenging task due to the fact that both phenomena affect the remitted or transmitted signals. Therefore, to unambiguously estimate absorption and scattering, techniques other than conventional wide-field imaging become imperative. Optical properties can be measured with temporally and spatially resolved techniques. The temporal and spatial domain can be divided into real and frequency domain via Fourier transforms (Fig. 1-4) [21].

Real temporal domain techniques measure the temporal point spread function of

tissues (t-$PSF(t)$, where t is time) by sending picosecond laser pulses via an optical fiber. The attenuation and broadening of the pulses are then recorded over time on the scale of nanoseconds [22, 4].

Temporal frequency domain techniques typically use an optical fiber to deliver light into the tissue, of which the intensity is modulated at hundreds of megahertz. The measured signal is the temporal modulation transfer function of tissues with respect to frequency (t-$MTF(\omega)$, where ω denotes temporal frequency). Absorption and scattering information can be deduced from the detected amplitude attenuation and the phase shift [4].

In contrast to temporally resolved approaches, spatially resolved techniques use continuous waves. Real spatial domain techniques typically illuminate tissues with pencil beams and measure tissue response using optical fibers at various source-detector separations [23]. By recording diffuse reflectance at multiple spatial locations of tissue, these techniques characterize the spatial point spread function (s-$PSF(\rho)$, where ρ is source-detector separation).

Although these three types of methods have different trade-offs in terms of hardware and model-fitting constraints [21], they share some common disadvantages. For example, they all require specialized apparatus that requires patient contact. Moreover, these probe-based devices are only capable of point measurements, making real-time, wide-field mapping of heterogeneous optical properties particularly challenging.

In recent years, a new technique emerged that measures the optical properties in the fourth domain, called Spatial Frequency Domain Imaging (SFDI). Compared

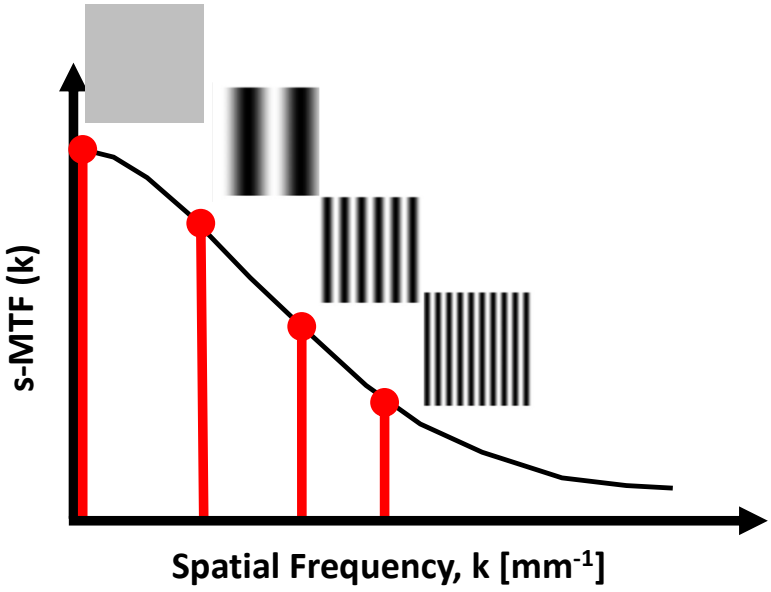

Figure 1-5. SFDI samples different points on the s-MTF curve by varying the spatial frequencies projected.

to the conventional methods, SFDI is a wide-field mapping approach that is easy to implement, and it will be the main focus of this dissertation.

1.2.2 Spatial Frequency Domain Imaging (SFDI)

Spatial Frequency Domain Imaging (SFDI) decouples absorption from scattering by characterizing the tissue modulation transfer function ($s\text{-}MTF(k)$) to spatially modulated light with spatial frequencies k [21, 24]. By projecting sinusoidal patterns at various spatial frequencies, SFDI samples different points on the $s\text{-}MTF(k)$ curve. In Fig. 1-5, the spatial frequency increases from left to right. 0mm^{-1} spatial frequency is simply a planar wave, and is often referred to as "DC", whereas non-zero spatial frequencies at spatially varying and called "AC".

The optical properties can be estimated from tissue spatial frequency response with either a diffusion-based or a transport-based approach [25]. The first approach relies on the diffusion approximation to the radiative transport equation. The time-independent form of the diffusion equation is:

$$\nabla^2\varphi - \mu_{eff}^2\varphi = -3\mu_{tr}q, \tag{1.8}$$

where φ is the fluence, $\mu_{eff} = (3\mu_a\mu_{tr})^{1/2}$ is the effective attenuation coefficient, $\mu_{tr} = (\mu_a + \mu_s')$ is the transport coefficient, and q is an isotropic source term (radiant energy density). Note that the optical penetration depth, δ_{eff}, is defined as the inverse of effective attenuation coefficient:

$$\delta_{eff} = \frac{1}{\mu_{eff}} \tag{1.9}$$

Assuming tissue response is radially symmetric and independent of the direction of modulation, we can express a periodic plane wave as:

$$q = q_0(z)cos(kx + \Phi), \tag{1.10}$$

where $q_0(z)$ is the depth-dependent intensity of the illumination, k is the spatial frequency, x is the location, and Φ is the spatial phase. If we assume normally incident light and a homogeneous medium, the fluence will have the same spatial frequency

and phase as the illumination [26]:

$$\varphi = \varphi_0(z)cos(kx + \Phi). \tag{1.11}$$

Substituting Eqs. 1.10 and 1.11 into Eq. 1.8, we obtain:

$$\frac{d^2}{dz^2}\varphi_0(z) - \mu_{eff}'^2\varphi_0(z) = -3\mu_{tr}q_0(z), \tag{1.12}$$

where $\mu'_{eff} = (\mu^2_{eff} + k^2)^{1/2}$. According to Eq. 1.9, the effective penetration depth now becomes:

$$\delta'_{eff} = \frac{1}{\mu'_{eff}} = \frac{1}{(3\mu_a(\mu_a + \mu'_s) + k^2)^{1/2}}. \tag{1.13}$$

Therefore, the penetration depth of spatially modulated light is related to the sample optical properties and the spatial frequency of the illumination. At zero spatial frequency ($k = 0$), δ'_{eff} reduces to δ_{eff}. For a set of optical properties, the $s\text{-}MTF(k)$ can then be characterized using the diffuse reflectance as a function of spatial frequency:

$$R_d(k) = \frac{3Aa'}{(\mu'_{eff}/\mu_{tr} + 1)(\mu'_{eff}/\mu_{tr} + 3A)}. \tag{1.14}$$

where

$$A = \frac{1 - R_{eff}}{2(1 + R_{eff})};$$

$$R_{eff} \approx 0.0636n + 0.668 + \frac{0.710}{n} + \frac{1.440}{n^2};$$

$$a' = \frac{\mu_s'}{\mu_{tr}}.$$

It is important to note that the diffusion approximation is only true for turbid media, where

$$\mu_s' \gg \mu_a, \tag{1.15}$$

and for relatively low spatial frequencies

$$f_x \ll \mu_{tr} = \frac{1}{l^*}. \tag{1.16}$$

Here, $f_x = (k_x/2\pi)$ is the transport spatial frequency (typically in mm^{-1}), and l^* is the transport mean free path.

With two or more spatial frequencies, we can measure the optical properties by obtaining an analytical solution to Eq. 1.14, which is intuitive and easy to implement. However, it does pose constraints on the spatial scales and optical properties [25]. Therefore, the transport-based approach is more commonly used, which simulates photon transport using Monte Carlo models [27, 28]. Typically, the spatially resolved diffuse reflectance ($R_d(\rho)$) is simulated, where ρ is the source-detector separation, and

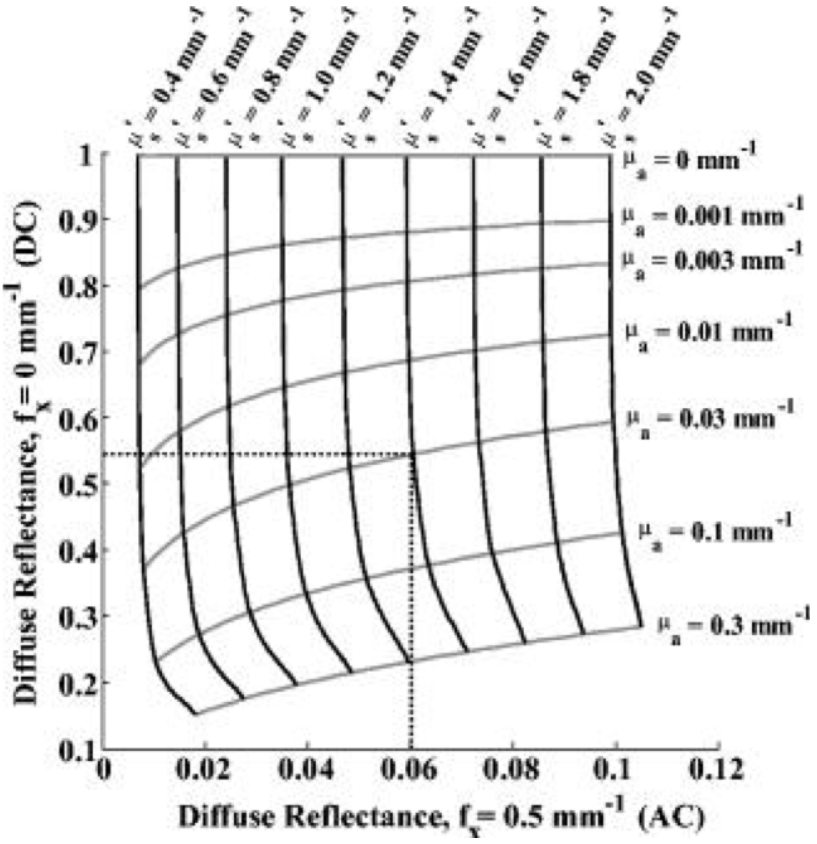

Figure 1-6. An example lookup table (LUT) that relates optical properties to spatial frequency-dependent diffuse reflectance. Figure adapted from [2].

$(R_d(k))$ can be computed using a 1-D Hankel transform of order zero:

$$R_d(k) = 2\pi \sum_{i=1}^{n} \rho_i J_0(k\rho_i) R_d(\rho_i) \Delta\rho_i. \tag{1.17}$$

J_0 is the zeroth-order Bessel function of the first kind, and $\Delta\rho_i$ stands for the interval of finite spatial bins. μ_a and μ_s' can then be estimated by searching a lookup table (LUT) that relates the optical properties to R_d at two or more distinct spatial frequencies (Fig. 1-6) [21].

The implementation of SFDI is fairly straightforward. Shown in Fig. 1-7, a typical SFDI setup includes a projector and camera. Cross-polarizers are often incorporated

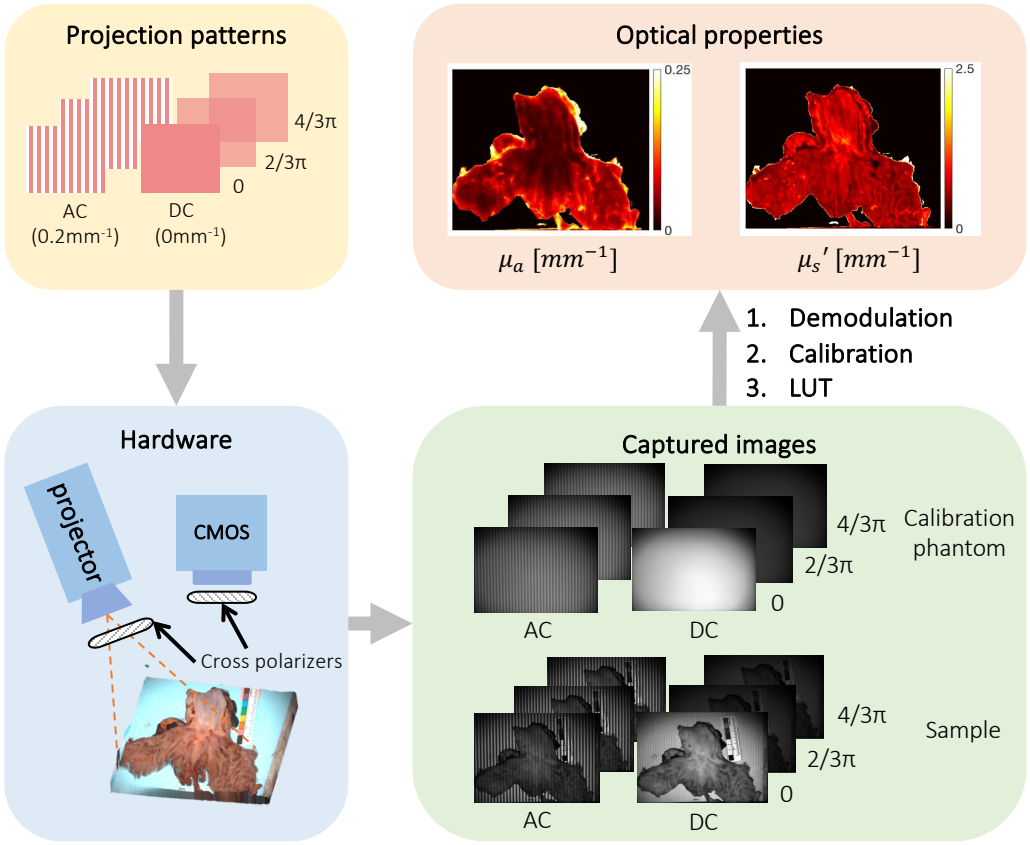

Figure 1-7. Overview of conventional SFDI illumination patterns, hardware, and processing flow. SFDI captures six frames (three phase offsets at two different spatial frequencies) to generate an absorption and reduced scattering map. DC indicates planar illumination images and AC indicates spatially modulated images. To calculate optical properties, acquired images are demodulated, calibrated against a reference phantom, and inverted using a lookup table. Figure adapted from [3].

to reduce the effect of specular reflection.

Although the most accurate results are achieved with multiple spatial frequencies, to balance the trade-off between acquisition speed and accuracy, conventional SFDI uses two spatial frequencies to decouple the effect of μ_a and μ_s'. In fact, SFDI using DC and a relatively low AC frequency, such as 0.1 or 0.2mm^{-1}, can achieve similar accuracy to using 13 spatial frequencies ranging from 0 to 0.5mm^{-1} [29]. In addition to faster acquisition, optical property inversion methods, such as LUT search that

relates optical properties to simulated spatial frequency-dependent diffuse reflectance, can be much more easily and efficiently implemented using two spatial frequencies with little accuracy loss. Thus, we selected 0 and 0.2mm^{-1} in this work.

At each spatial frequency, SFDI requires at least three phase offsets ($\Phi = 0$, $\frac{2}{3}\pi$, and $\frac{4}{3}\pi$), where:

$$I_i(x, f_x) = M_{AC}(x, f_x) * cos(2\pi f_x x + \Phi_i) + M_{DC}(x). \tag{1.18}$$

M_{AC} and M_{DC} stand for the demodulated AC and DC amplitude, respectively. M_{AC} can be computed from the three phase images using a demodulation equation:

$$M_{AC}(x) = \frac{\sqrt{2}}{3} \cdot \sqrt{(I_1(x) - I_2(x))^2 + (I_2(x) - I_3(x))^2 + (I_3(x) - I_1(x))^2}, \tag{1.19}$$

where I_1, I_2, and I_3 represent images at the three phase offsets. The purpose of this demodulation step is to extract the envelope of the spatially varying tissue response to a certain spatial frequency (Fig. 1-8).

Note that DC amplitude (M_{DC}) can be estimated as the average of the three AC images:

$$M_{DC}(x) = M_{AC}(x, f_x = 0) = \frac{1}{3}(I_1(x, f_x) + I_2(x, f_x) + I_3(x, f_x)). \tag{1.20}$$

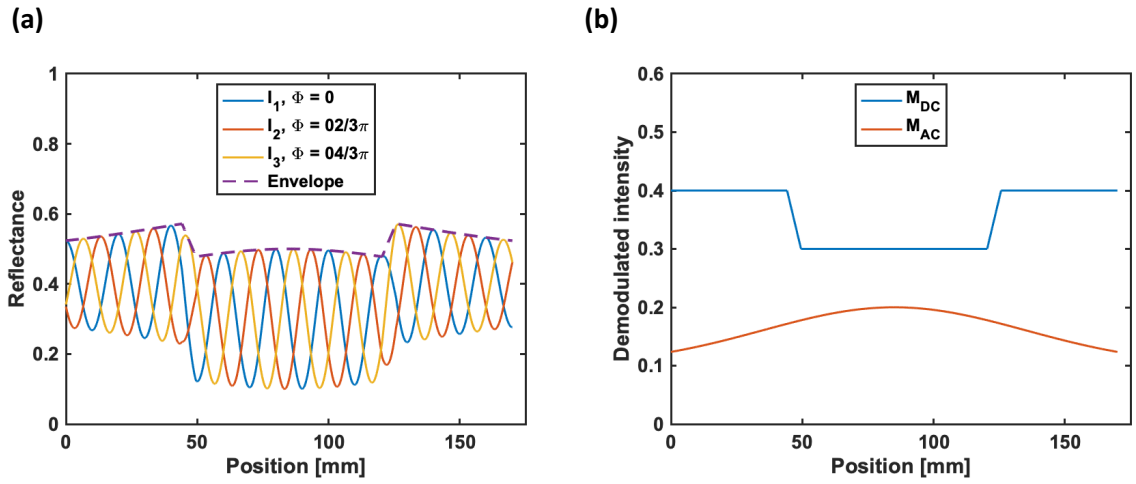

Figure 1-8. (a) Simulation results of modulated reflectance in one dimension of an SFDI image showing three phase offsets. The envelope is a sum of AC and DC amplitude. (b) Demodulated AC and DC amplitude.

However, unlike demodulation, Eq. 1.20 does not eliminate the effect of noise or ambient light. Certain types of noise, such as dark current, can fluctuate over time due to changes in temperature. Therefore, most researchers opt to capture three DC images and apply Eq. 1.19 to extract M_{DC}. In contrast to AC patterns, "phase shifts" at 0mm^{-1} are demonstrated as changes in the intensity of the planar waves. In this case, a minimum of six SFDI images is required for a sample of interest.

After demodulation, M_{AC} and M_{DC} of the sample of interest are calibrated against the response of a reference phantom with known optical properties (Fig. 1-7) to obtain the instrument-independent diffuse reflectance (R_d) at each pixel x:

$$R_d(x) = \frac{M_{AC}(x)}{M_{AC,ref}(x)} \cdot R_{d,predicted}. \tag{1.21}$$

Here, $M_{AC,ref}$ denotes the demodulated AC amplitude of the reference phantom,

and $R_{d,predicted}$ is the diffuse reflectance predicted by Monte Carlo models. From Eq. 1.21, it is apparent that a calibration phantom with known and spatially uniform optical properties is crucial for the accurate estimation of diffuse reflectance using SFDI.

Finally, the optical properties can be estimated by searching a lookup table (Fig. 1-6), which correlate $R_{d,AC}$ and $R_{d,DC}$ to μ_a and μ_s' at finite intervals. The resulting optical property maps have the same resolution as the original image.

So far, we have discussed the theory and implementation details of SFDI. A very important assumption here is that the sample of interest is flat with homogeneous surface profiles. However, this is not the case for most biological tissues, and SFDI without accounting for heterogeneous sample topography can produce errors greater than 100% [30]. Thus, profilometry correction is integral to achieving high accuracy with SFDI. Similar to SFDI structured illumination, surface profile measurements can be performed using sinusoidal illumination at different phase offsets and spatial frequencies [31–33]. Profile correction can then be performed from the measured sample shape, which includes height-dependent and angle-dependent intensity correction.

Height-dependent correction can be done via a physical inverse square model or calibration. Most studies use the calibration-based approach because it works on surfaces that are not Lambertian and can be easily adapted to different imaging geometries [30]. The modulation amplitude of the homogeneous reference phantom at

each pixel location is measured at n discrete heights:

$$M_{AC,ref,k}(f_x) = f(Height_{ref,k}), \tag{1.22}$$

where k ranges from 0 to n. The amplitude of the reference phantom is then corrected by interpolating the sample profilometry on a pixel basis using Eq. 1.22:

$$M_{AC,ref,height\text{-}corrected}(f_x) = f(Height_{sample}). \tag{1.23}$$

Angle correction is commonly done using a Lambertian reflectance model. The angles θ between surface normals of the sample at each pixel and the optical axis of the camera can be extracted from the measured profile, and the final profile corrected amplitude of the reference phantom is:

$$M_{AC,ref,profile\text{-}corrected}(f_x) = M_{AC,ref,height\text{-}corrected}(f_x) \times cos(\theta). \tag{1.24}$$

This Lambertian correction is valid for angles less than 40°. A recent study proposed a modified Lambertian correction that added a k exponential term to the original expression [34]:

$$M_{AC,ref,profile\text{-}corrected}(f_x) = M_{AC,ref,height\text{-}corrected}(f_x) \times cos(\theta)^k, \tag{1.25}$$

where $0 \leq \mathrm{k} \leq 1$. This method uses parameter optimization to find the best k for

AC and DC spatial frequency. It improves the accuracy of angle correction and is valid up to 75° angles. Interestingly, the study finds that the most accurate results are achieved without any angle correction at AC spatial frequency ($0.15mm^{-1}$).

Another important aspect of SFDI structured illumination is its ability to obtain optical section. Previously discussed in Eq. 1.13, the effective penetration depth of spatially modulated light is related to the spatial frequency and the optical properties. This is due to the fact that turbid media, such as biological tissues, act as a low-pass filter that limit the propagation of high spatial frequencies (Fig. 1-9). Note that in reality, photon penetration depth at a certain spatial frequencies follows a statistical distribution and can be more accurately modeled using Monte Carlo simulations [35]. For real tissue samples, although the demodulated signals are an aggregated effect of photons from a range of depths, we can clearly see more superficial tissue structures at higher spatial frequencies. Tissue microstructure imaging using high spatial frequencies has been an active area of research, and it has the potential to provide wide-field, slide-free pathology information [36, 37].

1.3 Applications and Limitations of SFDI

As discussed in previous sections, SFDI as an optical property measurement technique has many advantages compared to more conventional approaches in the first three domains (Fig. 1-4). For example, in contrast to most time-domain methods, typical SFDI setups can be easily implemented using consumer-grade components. SFDI is also capable of wide field-of-view optical property mapping instead of point measurements.

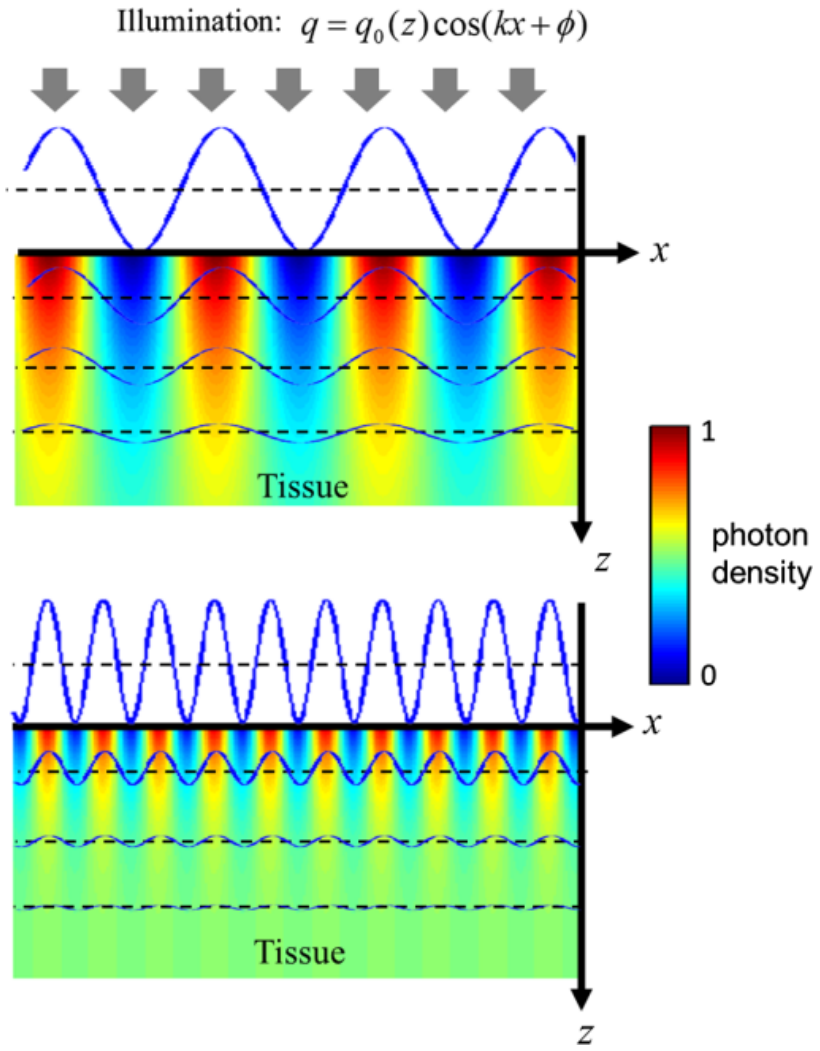

Figure 1-9. Turbid media, such as biological tissues, act as a low-pass filter. Higher spatial frequency light is attenuated more and thus penetrate less deep into the tissue. Figure adapted from [4].

Additionally, SFDI does not require patient contact, which is beneficial for patient monitoring and management.

Because of these advantages, SFDI has seen increasing use in many clinical applications. For example, high spatial-frequency SFDI is particularly useful for diagnostic assessment of breast lumpectomy samples. Shown in Fig. 1-11(a), demodulated SFDI images ("Optical Scatter") at high spatial frequencies show distinct contrast between

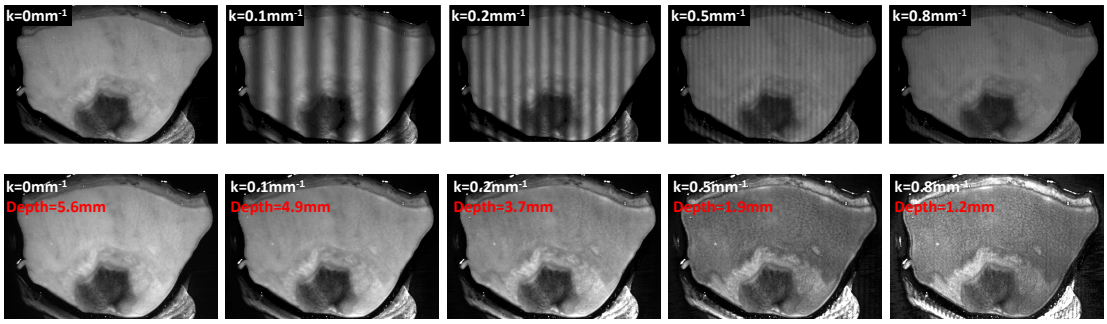

Figure 1-10. Example structured light (top row) and demodulated images (bottom row) of a pig colon stricture at various spatial frequencies. From left to right, as the spatial frequencies increase, superficial structures become more apparent in the demodulated images.

adipose (yellow arrow), connective (pink arrow), and malignant breast cancer tissue (red arrow). The same contrast is not seen in conventional white light images ("DWL Luminance") [5]. SFDI has also been approved by the FDA for diabetic foot ulcer staging applications. In Fig. 1-11(b), SFDI reveals a diabetic foot ulcer in the tissue oxygenation map that is not visible in the color image [6]. SFDI has been additionally used for image-guided surgery and patient monitoring during skin flap transplant procedures (Fig. 1-11(c)) [7].

Despite the advantages of SFDI, there are several technical challenges that limit its clinical adoption. For example, it requires six input images (three phase offsets at two spatial frequencies) to estimate the optical properties at one wavelength. To estimate chromophore concentrations and biomarkers such as tissue oxygenation, at least two wavelengths are needed, which results in twelve images per sample. This makes real-time monitoring difficult, and the accuracy could suffer from motion artifacts between consecutive frames. One potential way to circumvent this issue is to use Single Snapshot imaging of Optical Properties (SSOP) [38]. Discussed in more details

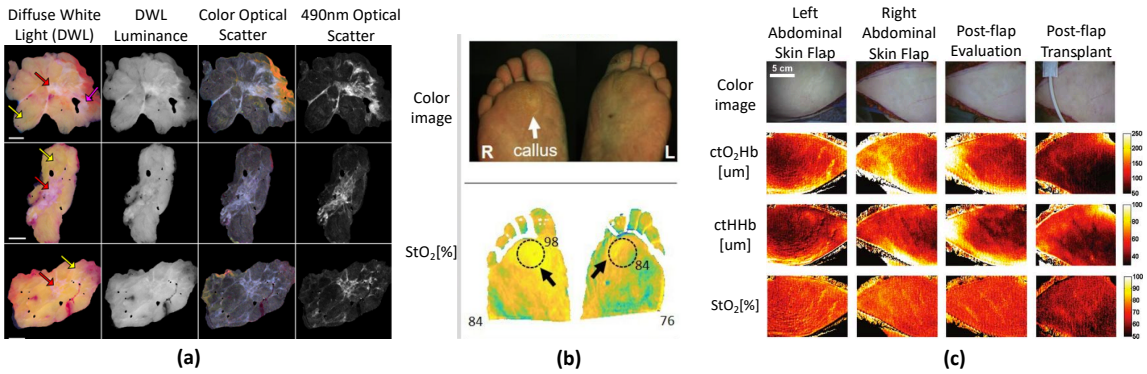

Figure 1-11. Applications of SFDI. (a) Breast lumpectomy margin assessment. Demodulated SFDI images ("Optical Scatter") at high spatial frequencies show distinct contrast between adipose (yellow arrow), connective (pink arrow), and malignant breast cancer tissue (red arrow). (b) Diabetic foot ulcer staging. SFDI reveals a diabetic foot ulcer in the tissue oxygenation map that is not visible in the color image. (c) Skin flap transplant monitoring. ctO_2Hb and CtHHB stand for the concentrations of oxygenated and de-oxygenated hemoglobin, respectively. Figures adapted from [5–7].

in later chapters, this technique uses frequency domain filtering to separate the AC and DC component in a single structured light image. Although this method makes real-time acquisition possible, it introduces errors and image artifacts in regions with low signal-to-noise ratios (SNR) or where the actual spatial frequency deviates from the expected one due to nonuniform surface topography and changes in working distance.

Another issue with SFDI is the use of a projector, which is difficult to incorporate into space-constrained clinical applications, such as flexible endoscopes. Furthermore, SFDI requires known spatial frequencies and carefully-controlled imaging geometries, which could be challenging in clinical settings.

1.4 Data-driven Methods for Quantitative Tissue Imaging

1.4.1 Convolutional Neural Networks

One of the objectives of this thesis is to design data-driven and machine learning methods for optical biomarker measurements. Specifically, we investigate the use of convolutional neural networks (CNN), which is a branch of machine learning that has seen rapid growth in the past decade for image-driven pattern recognition tasks [39, 40].

In this section, we will discuss the basics of artificial and convolutional neural networks. The CNN is a form of the artificial neural network (ANN), which resembles the biological nervous systems and consists of a large number of interconnected neurons (or nodes) [39]. An example of a simple ANN architecture is shown in Fig. 1-12. The network is a feed-forward model containing no cycles, which is commonly known as a multi-layer perceptron (MLP). Parameters in each perceptron or neuron, including weights and biases, are tuned and updated through the process of learning (Fig. 1-13). Input to the perceptron is first multiplied by its respective weight. The products from multiple input nodes are then summed and a bias is added. The result is passed through an activation function, the output of which becomes the input to the next layer in the MLP.

To update weights and biases, during each training iteration, the error of the network output is calculated using a loss function. The contribution of each weight

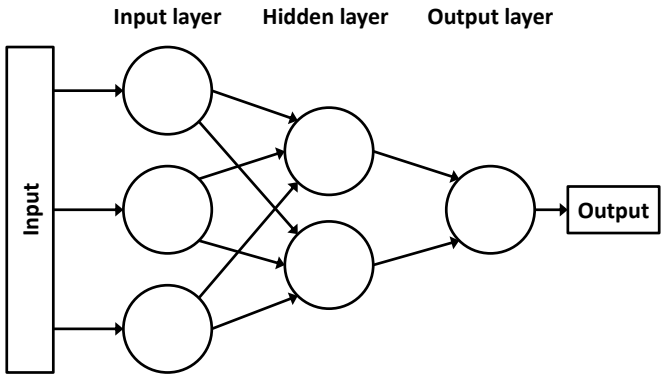

Figure 1-12. Example architecture of an artificial neural network (ANN), which is a feed-forward model commonly known as a multi-layer perceptron (MLP). It consists of an input layer, at least one hidden layer, and an output layer. Perceptrons (or neurons, represented as circles in this figure) are interconnected between layers so that output from a node in one layer becomes the input to all the nodes in the next layer. Parameters in each node are automatically tuned through the process of learning to optimize the output using information extracted from input.

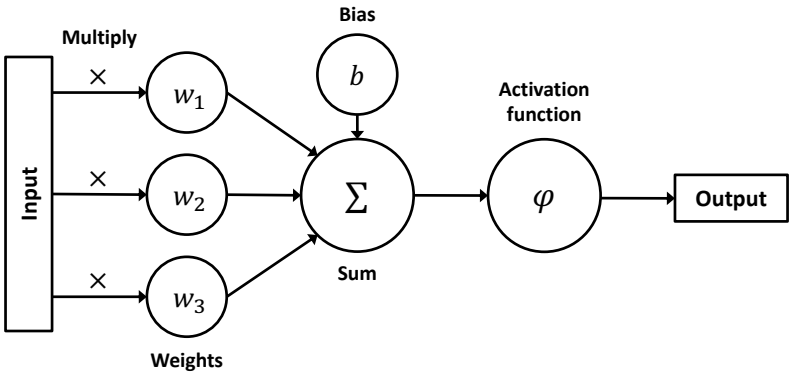

Figure 1-13. Structure of a perceptron (or neuron) in a multi-layer perceptron network. Input to the perceptron is first multiplied by a weight and summed across input nodes. A bias is added and the result is then passed through an activation function. The perceptron output becomes the input to the next layer. Weights and biases are updated using backpropagation and gradient descent.

and bias to the overall error is then determined using backpropagation [41], which computes the error gradients with respect to individual parameters. The parameters are then updated by taking steps in the opposite direction of the computed gradients in a process called "gradient descent", which identifies local minima of a differentiable loss function [42]. The network parameters optimized through the training process are used for testing or inference on data not included in the training set.

Another important aspect of the ANN is the activation function. Some commonly used activation functions are shown in Fig. 1-14. The purpose of an activation function is to determine the extent to which a neuron is activated. Moreover, in contrast to weights and biases that are only capable of linear operations, activation functions add non-linearity to the network so that it can learn more complex information. Note that ReLU and leaky ReLU are typically used in hidden layers to prevent the issue of vanishing gradients where training becomes stagnant [43]. This is because non-linear functions, such as sigmoid and Tanh, clamps the output maximum and can easily saturate during training.

The most important aspect that differentiates a CNN from conventional ANNs is the convolution operation, which is capable of leveraging spatial correlation among image pixels, making CNNs more suitable for image-driven applications. In image processing, the purpose of convolution is to transform image data using two-dimensional kernels or filters. Similar to ANN parameters, values of these convolutional kernels are learned and optimized in a CNN to produce the desired output. The size and number of kernels in each convolutional layers are network-specific design parameters

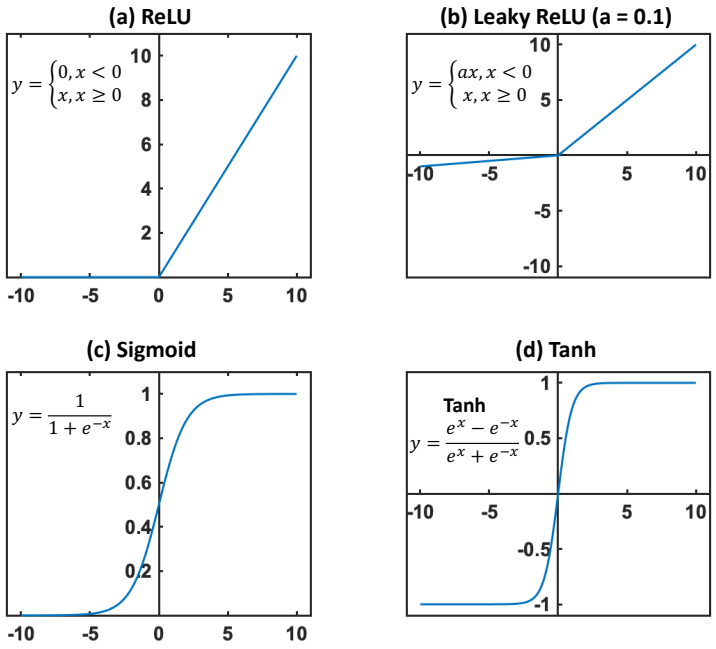

Figure 1-14. Commonly used activation functions in artificial neural networks. (a) ReLu; (b) Leaky ReLU with scale factor a=0.1; (c) Sigmoid; (d)Tanh. Activation functions determine if a neuron should be activated and introduce non-linearity to neural networks. ReLU and leaky ReLU are typically used in hidden layers to prevent vanishing gradients.

(Fig. 1-15). Padding may be necessary to retain the correct image dimensions.

After each convolutional layer, normalization is applied to the results to stabilize training by preventing the issue of exploding gradients where big error gradients accumulate and cause extremely large updates. Some commonly used normalization techniques include layer normalization, which normalizes across all output from a convolutional layer, and batch normalization, which normalizes across a mini-batch of training samples. When the size of the mini-batch is one, batch normalization is equivalent to instance normalization that is applied to output from a single sample.

Following normalization, an activation function is applied, and the output activation map is often referred to as a "feature map". Subsequently, a pooling layer is employed

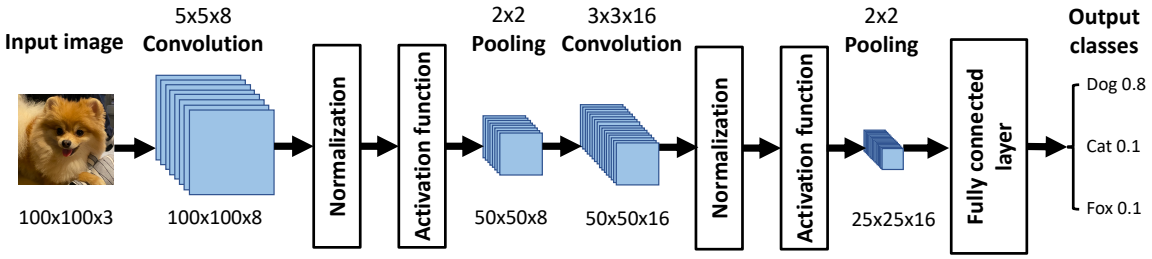

Figure 1-15. A simple convolutional neural network for image classification. The size and number of kernels can be different for each convolutional layers. Normalization is applied to the convolution results to stabilize training. An activation function then produces a feature map, which is typically followed by a pooling layer for dimensional reduction. A fully connected layer similar to a multi-layer perceptron is included for final class predictions.

to perform dimensional reduction that shrinks the feature map by the size of the pooling kernels. Two commonly used pooling operations are max and average pooling. The purpose of dimensional reduction via pooling is to reduce the computational cost by decreasing the number of parameters to train. It also allows the network to learn higher-level information by passing a more concise version of the feature map to the next convolutional layer.

The final component of a classification CNN is a fully connected layer, which is similar to a multi-layer perceptron. The output of this layer is passed through a non-linear activation function to produce a multinomial probability distribution of the input image belonging to each class, and the one with the highest score becomes the final class prediction.

In this thesis, we aim to use CNNs to generate optical biomarker maps from structured illumination images. In contrast to the classification network shown in Fig. 1-15, the output of our application is a two-dimensional image instead of a

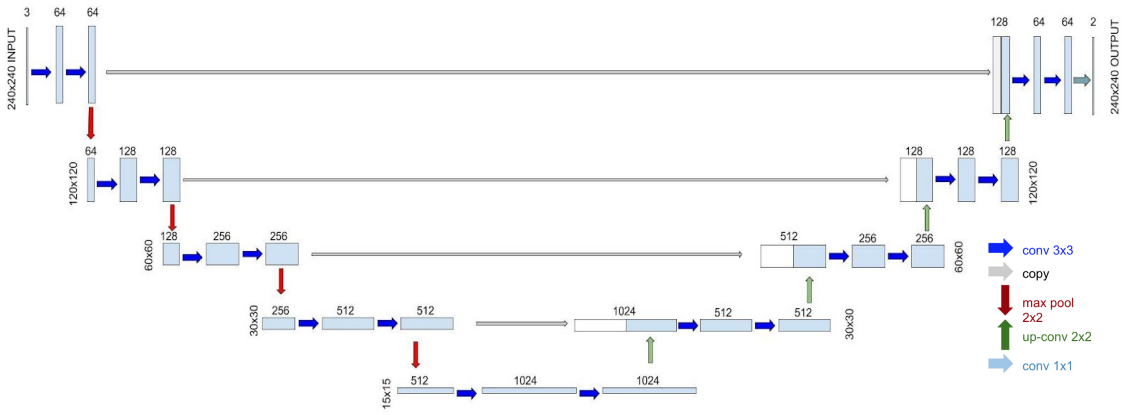

Figure 1-16. Example encoder-decoder model, called a U-Net, where encoder on the left side compresses input and maps it to a low-dimensional encoded space, and the decoder on the right side expands encoded data to produce output images. Skip connections between encoder and decoder components on the same level are employed to retain image resolution and improve convergence of deep networks with many layers. In the figure legend, "conv" denotes convolution and "up-conv" represents transposed convolution for upsampling. Horizontal numbers in the figure are the number of feature channels and vertical numbers are the dimensions of images or feature maps. The bottleneck portion in the bottom of the architecture has the smallest image dimensions and the largest number of feature maps.

single number representing the class label. To perform image generation, the fully connected layer in a classification network is replaced with another branch that up-samples the extracted features. This type of framework is often called an encoder-decoder model, where the encoder compresses input images into a low-dimensional latent space (or bottleneck) and the decoder expands and converts encoded data into desired output. An popular encoder-decoder framework, called a U-Net [44], is shown in Fig. 1-16. A representative feature of the U-Net is the long skip connections between encoder and decoder on the same level, which are employed to preserve image resolution and stabilize gradient updates in deep neural nets. Note that the image dimensions (vertical numbers in Fig. 1-16) decrease while the number of feature channels (horizontal numbers) increases as the model approaches the bottleneck.

1.4.2 Prior Work in Machine Learning for Quantitative Biomarker Mapping

In this section, we discuss the prior work in machine learning for optical property and chromophore imaging, especially in the context of SFDI as a wide-field and non-contact technique.

Panigrahi *et al.* [45] attempted to incorporate random forest regression as an inversion method to rapidly and accurately estimate optical properties from SFDI diffuse reflectance measurements. This technique used conventional three-phase-shift and two-spatial-frequency SFDI to obtain diffuse reflectance prior to applying the random forest model. Zhao *et al.* implemented an ultrafast deep learning framework with a multi-layer perceptron to predict highly-accurate optical properties from diffuse reflectance at five spatial frequencies [46]. The group later expanded upon this work and developed a deep residual network to predict the concentrations of oxygenated and de-oxygenated hemoglobin from diffuse reflectance at five spatial frequencies and two wavelengths [47]. Diffuse reflectance was obtained with conventional SFDI.

To balance the trade-off between acquisition speed and accuracy of SFDI, our group was the first to develop an end-to-end approach to directly estimate optical properties from single snapshot structured light images [3]. We posed this challenge as an image-to-image translation task and designed a conditional generative adversarial network (cGAN) for high-fidelity optical property mapping. Aguénounon *et al.* later achieved real-time estimation with profile correction using a custom-made GPU implementation of a twin U-Net framework [48]. Recently, our group leveraged the computational

power of NVIDIA TensorRT for reduced latency and optimized inference of trained models. We demonstrated real-time acquisition and processing of high-quality tissue oxygenation maps in an end-to-end fashion using a cGAN paradigm [49]. Overall, it can be concluded that machine learning is a powerful tool for improving the speed, accuracy, and image quality of SFDI as a powerful quantitative imaging technique.

1.5 Dissertation Outline and Overview

1.5.1 Thesis Statement

The thesis of this work is that ***data-driven methods can improve the performance of single-shot Spatial Frequency Domain Imaging (SFDI), and the use of laser speckles as structured light can improve the clinical adoptability of SFDI as an optical property mapping technique.*** The work designs and applies a novel deep learning framework and benchmarks against a variety of state-of-the-art methods. It also evaluates the performance of different single-shot SFDI techniques in terms of accuracy, image quality, and processing speed for both uncorrected and profile-corrected optical property and tissue biomarker measurements. Moreover, this work examines the use of random speckle patterns for optical property measurements and its potential to replace conventional projector-based sinusoidal illumination. Finally, this work incorporates laser illumination and machine learning models into a clinical endoscope for rapid and accurate tissue biomarker mapping.

1.5.2 Aim 1: Single-shot Tissue Optical Property and Biomarker Measurements Enabled by Machine Learning

The goal of this part of the thesis is to tackle the first challenge of SFDI clinical adoptions, which is the trade-off between speed and accuracy. In Chapters 2-3, a deep learning-based technique using convolutional neural networks is presented for single-shot optical property and tissue oxygenation predictions. The proposed technique is benchmarked against various learning-based and physics-based approaches for accuracy and image quality evaluation. Moreover, both DC and AC images are explored for estimating profile-corrected and uncorrected biomarkers. We highlight the importance of structured illumination for accurate measurements, especially in cases where different combinations of optical properties can result in the same DC reflectance. Conventional, six-image-per-wavelength SFDI is used as ground truth for supervised training, and the training and testing sets include a wide range of homogeneous phantoms, *ex vivo*, and *in vivo* swine and human specimens. We additionally study the potential of the technique for real-time biomarker measurements. Overall, the proposed method demonstrates high accuracy and generalizability when compared to other benchmarks. The work in Chapters 2-3 was presented at the following conferences:

- M. T. Chen and N. J. Durr, "Quantitative tissue property measurements with structured illumination and deep learning," in *2020 IEEE Photonics Conference (IPC)*. IEEE, pp. 1–2

- M. T. Chen and N. J. Durr, "Real-time oxygenation mapping from structured light imaging with deep learning," in *Optical Tomography and Spectroscopy of Tissue XIV*, vol. 11639. International Society for Optics and Photonics, 2021, p. 116391D

and published in the following journals:

- M. T. Chen, F. Mahmood, J. A. Sweer, and N. J. Durr, "Ganpop: Generative adversarial network prediction of optical properties from single snapshot wide-field images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1988–1999, 2020

- M. T. Chen and N. J. Durr, "Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning," *Journal of Biomedical Optics*, vol. 25, no. 11, p. 112907, 2020

1.5.3 Aim 2: Endoscopic Imaging of Tissue Biomarkers Using Speckle Illumination Spatial Frequency Domain Imaging

This part of the thesis aims to incorporate SFDI into space-constrained clinical applications. In Chapters 4-5, we explore endoscopic imaging of tissue optical properties using speckle illumination. In Chapter 4, we first present a signal processing-based approach for optical property measurements using random laser speckles as structured illumination. We apply this approach using homogeneous tissue-mimicking phantoms

and heterogeneous tissue samples. Conventional, profile-corrected SFDI is used as ground truth for performance evaluation. We investigate the effect of the number of speckle patterns, surface topography of the samples, and blood flow on the accuracy of the proposed method.

In Chapter 5, the approach described in Chapter 4 is incorporated into a clinical endoscope. We use stereo vision to estimate sample profilometry and machine learning to improve the speed and accuracy of optical property predictions with speckle illumination. We demonstrate the hardware setup and some preliminary results, and we additionally discuss current limitations and future steps for this project. The work in Chapters 4-5 was presented at the following conference:

- M. T. Chen, T. L. Bobrow, and N. J. Durr, "Towards sfdi endoscopy with structured illumination from randomized speckle patterns," in *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XIX*, vol. 11631. International Society for Optics and Photonics, 2021, p. 116310Y

and published in the following journal:

- M. T. Chen, M. Papadakis, and N. J. Durr, "Speckle illumination sfdi for projector-free optical property mapping," *Optics letters*, vol. 46, no. 3, pp. 673–676, 2021

In Chapter 6, the key findings and contributions of this work are reviewed, and the current limitations and future directions are outlined. Moreover, we discuss the

implications for ongoing research on the utility of SFDI for wide-field diagnostic assessments of gastrointestinal tissues.

Chapter 2

Generative Adversarial Network Prediction of Optical Properties from Single Snapshot Wide-field Images

This chapter contains work originally published in [3]:

2.1 Introduction

In Chapter 1, the importance of tissue optical properties, including the absorption (μ_a) and reduced scattering (μ_s') coefficients, was discussed. Tissue optical properties can be useful clinical biomarkers for measuring trends and detecting abnormalities in tissue metabolism, tissue oxygenation, and cellular proliferation [13–17]. Optical properties can also be used for contrast in functional or structural imaging [19, 20].

To measure optical properties, it is generally necessary to decouple the effects of scattering and absorption, which both influence the measured intensity of remitted light. Separation of these parameters can be achieved with temporally or spatially resolved techniques, which can each be performed with measurements in the real or frequency domains. Spatial Frequency Domain Imaging (SFDI) decouples absorption from scattering by characterizing the tissue modulation transfer function to spatially modulated light [24, 21]. This approach has significant advantages in that it can easily be implemented with a consumer grade camera and projector, and achieve rapid, non-contact mapping of optical properties. These advantages make SFDI well-suited for applications that benefit from wide-field characterization of tissues, such as image-guided surgery [54, 7], tissue perfusion measurement [55], and wound characterization [56, 57]. Additionally, recent work has explored the use of SFDI for improving endoscopic procedures [58, 59].

Although SFDI is finding a growing number of clinical applications, there are remaining technical challenges that limit its adoption. First, SFDI requires structured light projection with carefully-controlled working distance and calibration, which

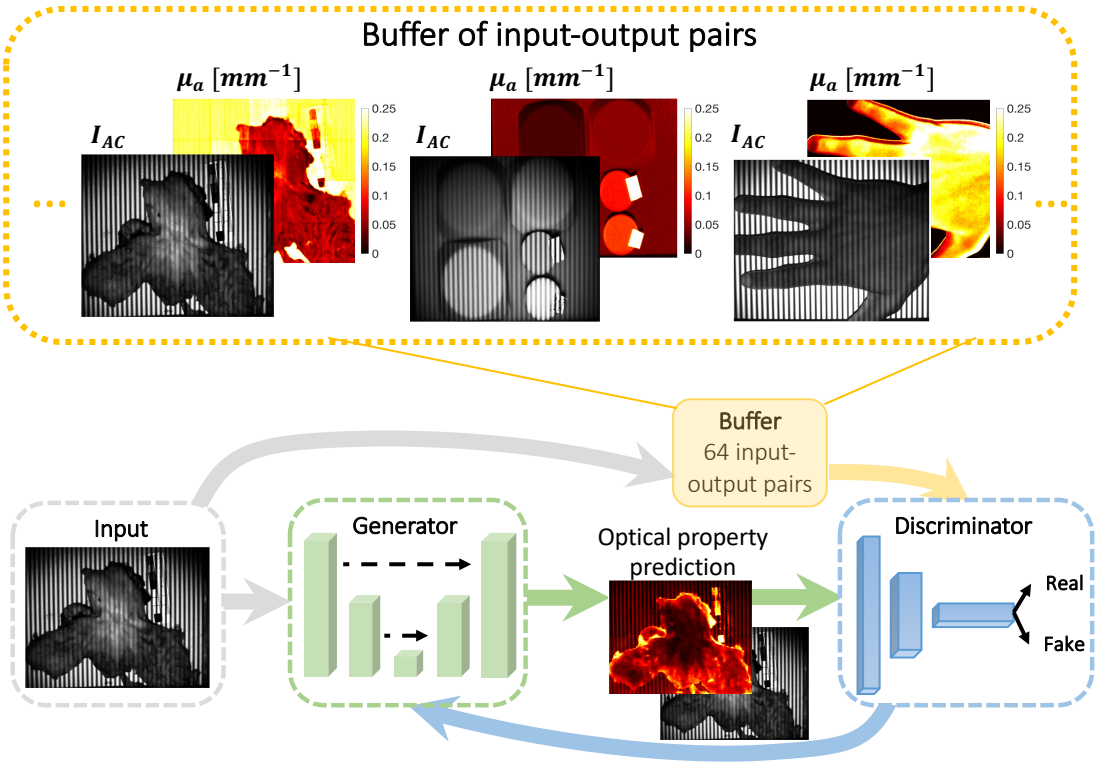

Figure 2-1. Proposed conditional Generative Adversarial Network (cGAN) architecture. The generator is a combination of ResNet and U-Net and is trained to produce optical property maps that closely resemble SFDI output. The discriminator is a three-layer classifier that operates on a patch level and is trained to classify the output of the generator as ground truth (real) or generated (fake). The discriminator is updated using a history of 64 previously-generated image pairs.

is especially challenging in an endoscopic setting. Second, it is difficult to achieve real-time measurements. Conventional SFDI requires a minimum of six images per wavelength (three distinct spatial phases at two spatial frequencies) to generate a single optical property map. A lookup table (LUT) search is then performed for optical property fitting. The recent development of real-time single snapshot imaging of optical properties (SSOP) has reduced the number of images required per wavelength from 6 to 1, considerably shortening acquisition time [38]. However, SSOP introduces image artifacts arising from single-phase projection and frequency filtering, which

corrupt the optical property estimations. To reduce barriers to clinical translation, there is a need for optical property mapping approaches that are simultaneously fast and accurate while requiring minimal modifications to conventional camera systems.

Here, we introduce a deep learning framework to predict optical properties directly from single images. Deep networks, especially convolutional neural networks (CNNs), are growing in popularity for medical imaging tasks, including computer-aided detection, segmentation, and image analysis[60–62]. We pose the optical property estimation challenge as an image-to-image translation task and employ generative adversarial networks (GANs) to efficiently learn a transformation that is robust to input variety. First proposed in [63], GANs have improved upon the performance of CNNs in image generation by including both a generator and a discriminator. The former is trained to produce realistic output, while the latter is tasked to classify generator output as real or fake. The two components are trained simultaneously to outperform each other, and the discriminator is discarded once the generator has been trained. When both components observe the same type of data, such as text labels or input images, the GAN model becomes conditional. Conditional GANs (cGANs) are capable of making structured predictions by incorporating non-local, high-level information. Moreover, because they can automatically learn a loss function instead of using a handcrafted one, cGANs have the potential to be an effective and generalizable solution to various image-to-image translation tasks [64, 65]. In medical imaging, cGANs have been proven successful in many applications, such as image synthesis [66], noise reduction [67], and sparse reconstruction [68]. In this study, we train cGAN

networks on a series of structured (AC GANPOP) or flat-field illumination images (DC GANPOP) paired with corresponding optical property maps (Fig. 2-1). We demonstrate that the GANPOP approach produces rapid and accurate estimation from input images from a wide variety of tissues using a relatively small set of training data.

2.2 Related Work

2.2.1 Diffuse Reflectance Imaging

Optical absorption and reduced scattering coefficients can be measured using temporally or spatially resolved diffuse reflectance imaging. Approaches that rely on point illumination inherently have a limited field of view [69, 70]. Non-contact, hyperspectral imaging techniques measure the attenuation of light at different wavelengths, from which the concentrations of tissue chromophores, such as oxy- and deoxy-hemoglobin, water, and lipids, can be quantified [71]. A recent study has also proposed using a Bayesian framework to infer tissue oxygen concentration by recovering intrinsic multispectral measurements from RGB images [72]. However, these methods fail to unambiguously separate absorption and scattering coefficients, which poses a challenge for precise chromophore measurements. Moreover, accurate determination of both parameters is critical for the detection and diagnosis of diseases [13, 17].

2.2.2 Single Snapshot Imaging of Optical Properties

SSOP achieves optical property mapping from a single structured light image. Using Fourier domain filtering, this method separates DC (planar) and AC (spatially modulated) components from a single-phase structured illumination image [38]. A grid pattern can also be applied to simultaneously extract optical properties and three-dimensional profile measurements [73]. When tested on homogeneous tissue-mimicking phantoms, this method is able to recover optical properties within 12% for absorption and 6% for reduced scattering using conventional profilometry-corrected SFDI as ground truth.

2.2.3 Machine Learning in Optical Property Estimation

Despite its prevalence and increasing importance in the field of medical imaging, machine learning has only recently been explored for optical property mapping. This includes a random forest regressor to replace the nonlinear model inversion [74], and using deep neural networks to reconstruct optical properties from multifrequency measurements [46]. Both of these approaches aim to bypass the time-consuming LUT step in SFDI. However, they require diffuse reflectance measurements from multiple images to achieve accurate results and process each pixel independently without considering the content of surrounding pixels.

2.3 Contributions

We propose an adversarial framework for learning a content-aware transformation from single illumination images to optical property maps. In this work, we:

1.) develop a data-driven model to estimate optical properties directly from a single input reflectance image;

2.) demonstrate advantages of structured (AC) versus flat-field (DC) illumination to determine optical properties via an adversarial learning approach;

3.) perform cross-validated experiments, comparing our technique with model-based SSOP and other deep learning-based methods; and

4.) acquire and make publicly-available a dataset of registered flat-field-illumination images, structured-illumination images, and ground-truth optical properties of a variety of *ex vivo* and *in vivo* tissues.

2.4 Methods

For training and testing of the GANPOP model, single structured or flat-field illumination images were used, paired with registered optical property maps. Ground truth optical properties were obtained by conventional six-image SFDI. GANPOP performance was analyzed and compared to other techniques both in unseen tissues of the same type as the training data (human *ex vivo* esophagus images from a new patient) and in different tissue types (*in vivo* and *ex vivo* swine gastrointestinal tissues).

2.4.1 Hardware

In this study, all images were captured using a commercial SFDI system (Reflect RS^{TM}, Modulated Imaging Inc.). A schematic of the system is shown in Fig. 1-7. Cross polarizers were utilized to reduce the effect of specular reflections, and images were acquired in a custom-built light enclosure to minimize ambient light. Raw images, after 2×2 pixel hardware binning, were 520×696 pixels, with a pixels size of 0.278 mm in the object space.

2.4.2 SFDI Ground Truth Optical Properties

Ground truth optical property maps were generated using conventional SFDI with 660 nm light following the method from Cuccia *et al.* [21]. First, images of a calibration phantom with homogeneous optical properties and the tissue of interest are captured under spatially modulated light. We used a flat polydimethylsiloxane-titanium dioxide (PDMS-TiO_2) phantom with reduced scattering coefficient of 0.957 mm^{-1} and absorption coefficient of 0.0239 mm^{-1} at 660 nm. We project spatial frequencies of 0 mm^{-1} (DC) and 0.2 mm^{-1} (AC), each at three different phase offsets (0, $\frac{2}{3}\pi$, and $\frac{4}{3}\pi$) for this study. AC images are demodulated at each pixel using Eq. 1.19. The spatially varying DC amplitude is calculated as the average of the three DC images. Diffuse reflectance at each pixel is then computed as Eq. 1.21. Where indicated, we corrected for height and surface angle variation of each pixel from depth maps measured via profilometry. Profilometry measurements were obtained by projecting a spatial frequency of 0.15 mm^{-1} and calculating depth at each pixel [30].

A k value of 0.8 and 0.4 was used at DC and AC in order to minimize error in angle correction [34]. Finally, μ_a and μ_s' are estimated by fitting $R_{d,0mm^{-1}}$ and $R_{d,0.2mm^{-1}}$ into an LUT previously created using Monte Carlo simulations [75].

2.4.3 Single Snapshot Optical Properties (SSOP)

SSOP was implemented as the model-based alternative of AC GANPOP. This method separates DC and AC components from a single-phase structured-light image by frequency filtering with a 2D band-stop filter and a high-pass filter [38]. Both filters are rectangular windows that isolate the frequency range of interest while preserving high-frequency content of the image. In this study, cutoff frequencies f_{DC} = [0.16 mm^{-1}, 0.24 mm^{-1}] and f_{AC} = [0, 0.16 mm^{-1}] were selected [73]. M_{DC} can subsequently be recovered through a 2D inverse Fourier transform, and the AC component is obtained using an additional Hilbert transform.

2.4.4 GANPOP Architecture

The GANPOP architecture is based on an adversarial training framework. When used in a conditional GAN-based image-to-image translation setup, this framework has the ability to learn a loss function while avoiding the uncertainty inherent in using handcrafted loss functions [65, 76]. The generator is tasked with predicting pixel-wise optical properties from SFDI images while the discriminator classifies pairs of SFDI images and optical property maps as being real or fake (Fig. 2-1). The discriminator additionally gives feedback to the generator over the course of training. The generator

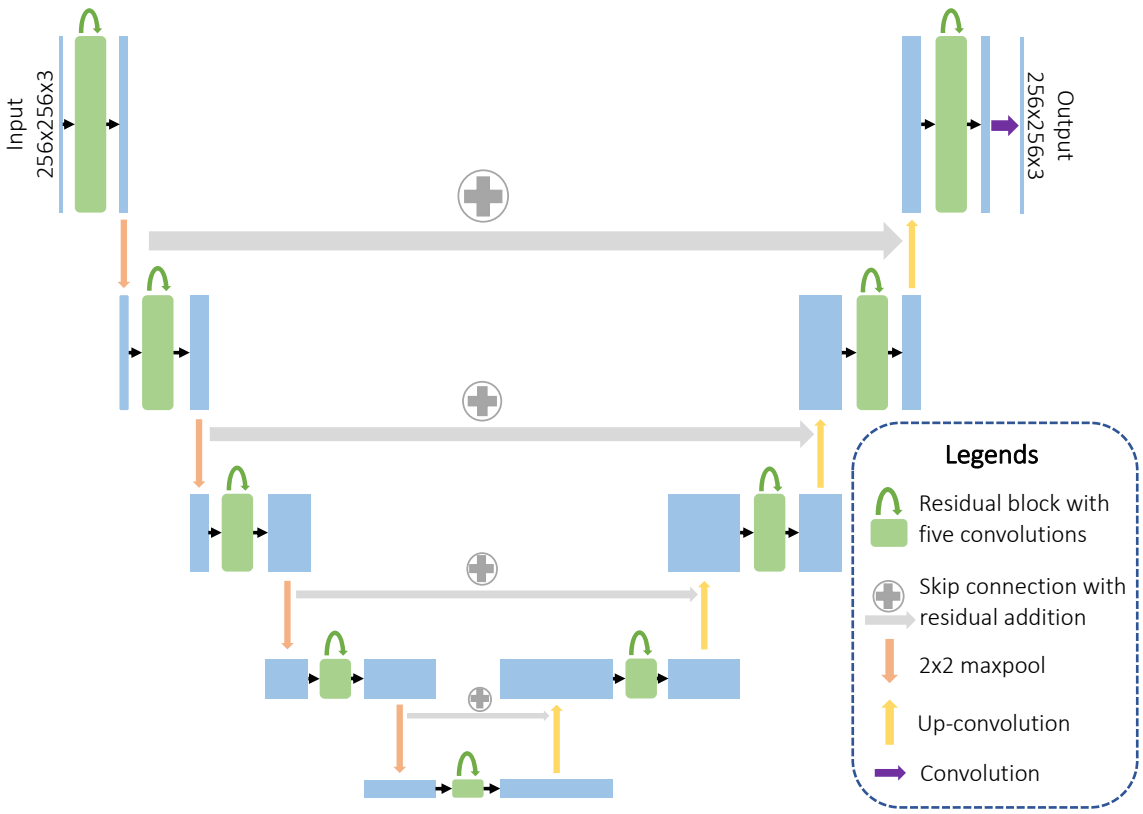

Figure 2-2. Detailed architecture of the proposed generator. We use a fusion network that combines properties from ResNet and U-Net, including both short and long skip connections in the form of feature addition. Each residual block contains five convolution layers, with short skips between the first and the fifth layer.

employs a modified U-Net consisting of an encoder and a decoder with skip connections [44]. However, unlike the original U-Net, the GANPOP network includes properties of a ResNet, including short skip connections within each level [77] (Fig. 2-2). Each residual block is a 3-layer building block with an additional convolutional layer on both sides. This ensures that the number of input features matches that of the residual block and that the network is symmetric [78]. Moreover, GANPOP generator replaces the U-Net concatenation step with feature addition, making it a fully residual network. Using n as the total number of layers in the encoder-decoder network and i as the current layer, long skip connections are added between the i^{th} and the $(n-i)^{th}$ layer

in order to sum features from the two levels. After the last layer in the decoder, a final convolution is applied to shrink the number of output channels and is followed by a *Tanh* function. Regular ReLUs are used for the decoder and leaky ReLUs (slope = 0.2) for the encoder. We chose a receptive field of 70 × 70 pixels for our discriminator because this window captures two periods of AC illumination in each direction. This discriminator is a three-layer classifier with leaky ReLUs (slope = 0.2), as discussed in [65]. The discriminator makes classification decisions based on the current batch as well as a batch randomly sampled from 64 previously generated image pairs. Both networks are trained iteratively and the training process is stabilized by incorporating spectral normalization in both the generator and the discriminator [79]. The conditional GAN objective for generating optical property maps from input images ($G : X \rightarrow Y$) is:

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D) =& \mathbb{E}_{x,y\sim p_{\text{data}}(x,y)}[(D(x, y) - 1)^2] \\ &+ \mathbb{E}_{x\sim p_{\text{data}}(x)}[D(x, G(x))^2],\end{aligned} \tag{2.1}$$

where G is the generator, D the discriminator, and p_{data} is the optimal distribution of the data. We empirically found that a least squares GAN (LSGAN) objective [80] produced slightly better performance in predicting optical properties than a traditional GAN objective [63], and so we utilize LSGAN in the networks presented here. An additional $\mathcal{L}_1$ loss term was added to the GAN loss to further minimize the distance

from the ground truth distribution and stabilize adversarial training:

$$\mathcal{L}_1(S) = \mathbb{E}_{x,y\sim p_{\text{data}}}(x,y)[||y - G(x)||_1]. \tag{2.2}$$

The full objective can be expressed as:

$$\mathcal{L}(G,D) = \mathcal{L}_{\text{GAN}}(G,D) + \lambda\mathcal{L}_1(G), \tag{2.3}$$

where λ is the regularization parameter of the $\mathcal{L}_1$ loss term. This optimization problem was solved using an *Adam* solver with a batch size of 1 [81]. The training code was implemented using Pytorch 1.0 on Ubuntu 16.04 with Google Cloud. A single NVIDIA Tesla P100 GPU was used for both training and testing. For all experiments, λ was set to 60. A total of 200 epochs was used with a learning rate of 0.0002 for half of the epochs and the learning rate was linearly decayed for the remaining half. Both networks were initialized from a Gaussian distribution with a mean and standard deviation of 0 and 0.02, respectively. In addition, we performed comparative analyses of our proposed network against other commonly-used architectures for image-to-image translation, including ResNet and U-Net, both standalone and incorporated into a GAN structure.

Conventional neural networks typically operate on three-channel (or RGB) images as input and output. In this study, four separate networks (N_1 to N_4) were trained for image-to-image translation with a variety of input and output parameters, summarized in Table 2-I.

Table 2-I. Summary of networks trained in this study.

	Input		Output	
N_i	Channel 1	Channel 2	Channel 1	Channel 2
N_1	$\frac{I_{AC}}{M_{DC,ref}}$	$\frac{I_{AC}}{M_{AC,ref}}$	μ_a	μ_s'
N_2	$\frac{I_{DC}}{M_{DC,ref}}$	$\frac{I_{DC}}{M_{AC,ref}}$	μ_a	μ_s'
N_3	$\frac{I_{AC}}{M_{DC,ref}}$	$\frac{I_{AC}}{M_{AC,ref}}$	$\mu_{a,corr}$	$\mu_{s,corr}'$
N_4	$\frac{I_{DC}}{M_{DC,ref}}$	$\frac{I_{DC}}{M_{AC,ref}}$	$\mu_{a,corr}$	$\mu_{s,corr}'$

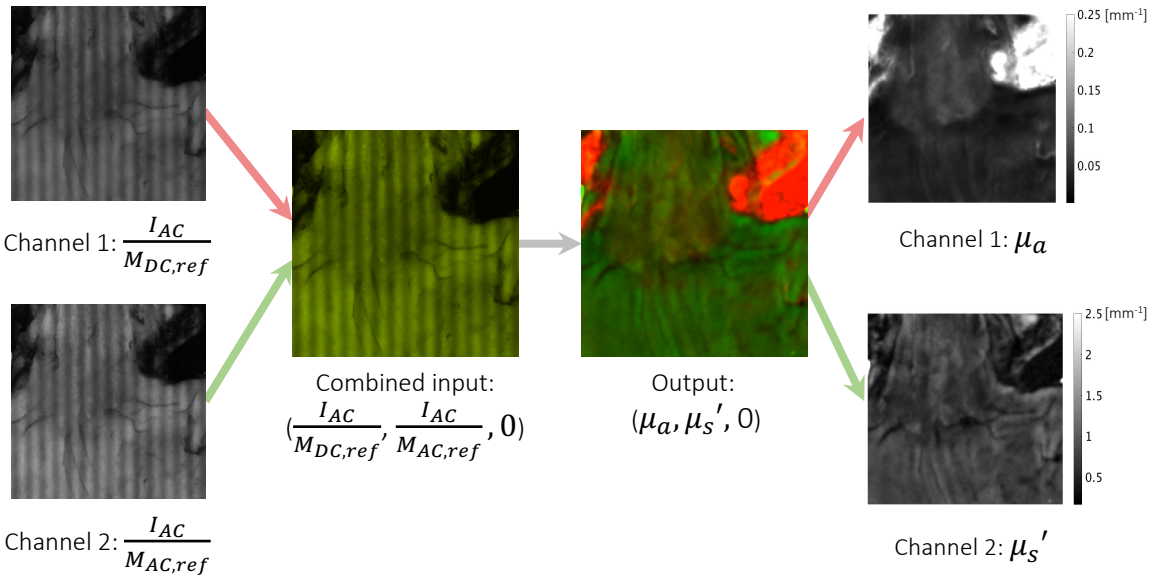

Figure 2-3. Example input-output pair used in N_1 showing each individual channel as well as the combined RGB images. Channel 1 and 2 of the output contain μ_a and μ_s', respectively. Thus, a high absorption appears red while a high scattering appears green.

For input, I_{AC} and I_{DC} represent single-phase raw images at 0.2 mm^{-1} and 0 spatial frequency, respectively. $M_{DC,ref}$ and $M_{AC,ref}$ are the demodulated DC and AC amplitude of the calibration phantom. Channel 3 is left as zeros in all cases. It is important to note that $M_{AC,ref}$ and $M_{DC,ref}$ are measured only once during calibration before the imaging experiment and thus do not add to the total acquisition time. The purpose of these two terms is to account for drift of the system over time and correct for non-uniform illumination, making the patch used in the network origin-independent. These two calibration images are also required by traditional SFDI and the SSOP approaches. A network without calibration was empirically trained, and it produced 230% and 58% larger error than with calibration in absorption and scattering coefficients, respectively. A single output image contains both μ_a and μ_s' in different channels. Two dedicated networks were empirically trained for estimating μ_a and μ_s' independently, but no accuracy benefits were observed. Optical property maps calculated by non-profile-corrected SFDI were used as ground truth for N_1 and N_2. We also assessed the ability of GANPOP networks (N_3 and N_4) to learn both optical property estimation and sample height and surface normal correction by training and testing with profilometry-corrected data ($\mu_{a,corr}$ and $\mu_{s,corr}'$).

All optical property maps for training and testing were normalized to have a consistent representation in the 8-bit images commonly used in CNNs. We defined the maximum value of 255 to be 0.25 mm^{-1} for μ_a and 2.5 mm^{-1} for μ_s'. Additionally, each image of size 520 $\times$ 696 was segmented at a random stride size into multiple patches of 256 $\times$ 256 pixels and paired with a registered optical property patch for

training, as shown in Fig. 2-3.

2.4.5 Tissue Samples

2.4.5.1 *Ex Vivo* Human Esophagus

Eight *ex vivo* human esophagectomy samples were imaged at Johns Hopkins Hospital for training and testing of our networks. All patients were diagnosed with esophageal adenocarcinoma and were scheduled for an esophagectomy. The research protocol was approved by the Johns Hopkins Institutional Review Board and consents were acquired from all patients prior to each study. All samples were handled by a trained pathologist and imaged within one hour after resection [82].

Example raw images of a specimen captured by the SFDI system are shown in Fig. 2-3, and 2-11(a). All samples consisted of the distal esophagus, the gastroesophageal junction, and the proximal stomach. The samples contain complex topography and a relatively wide range of optical properties (0.02-0.15 mm^{-1} for μ_a and 0.1-1.5 mm^{-1} for μ_s' at $\lambda = 660$ nm), making it suitable for training a generalizable model that can be applied to other tissues with non-uniform surface profiles. An illumination wavelength of 660 nm was chosen because it is close to the optimal wavelength for accurate tissue oxygenation measurements [1].

In this study, six *ex vivo* human esophagus samples were used for training of the GANPOP model and two used for testing. A leave-two-out cross validation method was implemented, resulting in four iterations of training for each network. Performance results reported here are from an average of these four iterations.

2.4.5.2 Homogeneous Phantoms

The four GANPOP networks were also trained on a set of tissue-mimicking silicone phantoms made from PDMS-TiO_2 (P4, Eager Plastics Inc.) mixed with India ink as absorbing agent [83]. To ensure homogeneous optical properties, the mixture was thoroughly combined and poured into a flat mold for curing. In total, 18 phantoms with unique combinations of μ_a and μ_s' were fabricated, and their optical properties are summarized in Fig. 2-4.

In this study, six tissue-mimicking phantoms were used for training and twelve for testing. We intentionally selected phantoms for training that had optical properties not spanned by the human training samples (highlighted by green ellipses in Fig. 2-5), in order to develop GANPOP networks capable of estimating the widest range of optical properties.

2.4.5.3 *In Vivo* Samples

To provide the network with *in vivo* samples that were perfused and oxygenated, eight human hands and feet with different levels of pigmentation (Fitzpatrick skin types 1-6) were imaged with SFDI. This protocol was approved by the Johns Hopkins Institutional Review Board and consent was acquired from each participant prior to imaging. Six hands and feet were used for training and two for testing. Similar to human esophagus samples, leave-two-out cross validations were used for each network.

2.4.5.4 Swine Tissue

Four specimens of upper gastrointestinal tracts that included stomach and esophagus were harvested from four different pigs for *ex vivo* imaging with SFDI. Optical properties of these samples are summarized in Fig. 2-6. Additionally, we imaged a pig colon *in vivo* during a surgery. The live study was performed with approval from Johns Hopkins University Animal Care and Use Committee (ACUC). All swine tissue images were excluded from training and used only for testing optical property prediction.

2.4.6 Performance Metric

Normalized Mean Absolute Error (NMAE) was used to evaluate the performance of different methods, which was calculated using:

$$NMAE = \frac{\sum_{i=1}^{T} |p_i - p_{i,ref}|}{\sum_{i=1}^{T} p_{i,ref}}. \tag{2.4}$$

p_i and $p_{i,ref}$ are pixel values of predicted and ground-truth data, and T is the total number of pixels. The metric was calculated using SFDI output as ground truth. A smaller NMAE value indicates better performance.

Additional metrics were used to assess pixel accuracy at constant reflectance values, including normalized error:

$$error_i = \frac{p_i - p_{i,ref}}{p_{i,ref}}, \tag{2.5}$$

and average optical property (OP) deviation:

$$\text{Average OP Deviation} = \frac{\sum_{i=1}^{T} \sqrt{(error_{\mu_a,i}^2 + error_{\mu_s',i}^2)}}{T}. \tag{2.6}$$

2.5 Results

2.5.1 SSOP Validation

For benchmarking, SSOP was implemented as a model-based counterpart of GANPOP. For independent validation, we applied SSOP to 18 homogeneous tissue phantoms (Fig. 2-4). Each value was calculated as the mean of a 100 × 100-pixel region of interest (ROI) from the center of the phantom, with error bars showing standard deviations. SSOP demonstrates high accuracy in predicting optical properties of the phantoms, with an average percentage error of 2.35% for absorption and 2.69% for reduced scattering.

2.5.2 AC GANPOP Test in Homogeneous Phantoms

Phantom optical properties predicted by N_1 are plotted with ground truth in Fig. 2-5. Each optical property reported is the average value of a 100 × 100 ROI of a homogeneous phantom, with error bars showing standard deviations. On average, AC GANPOP produced 4.50% error for absorption and 1.46% for scattering. The scatter plot in Fig. 2-5 is overlaid on a 2D histogram of pixel counts for each (μ_a, μ_s') pair used in an example training iteration. Green ellipses indicate training samples from

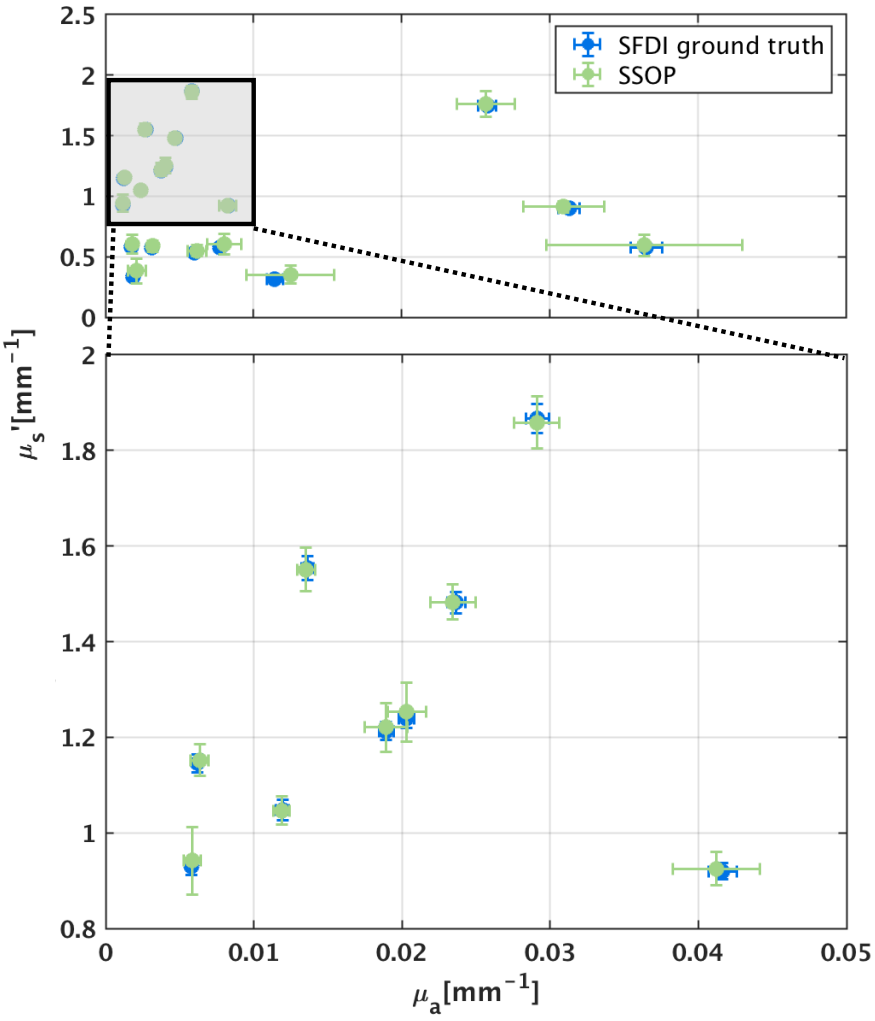

Figure 2-4. Validation of SSOP model-based prediction of optical properties with 18 homogeneous tissue phantoms. SSOP prediction from a single input image demonstrates close agreement with SFDI prediction from six input images.

homogeneous phantoms. The two testing results enclosed by red boxes have optical properties outside of the range spanned by the training data but were still reasonably estimated by the AC GANPOP network.

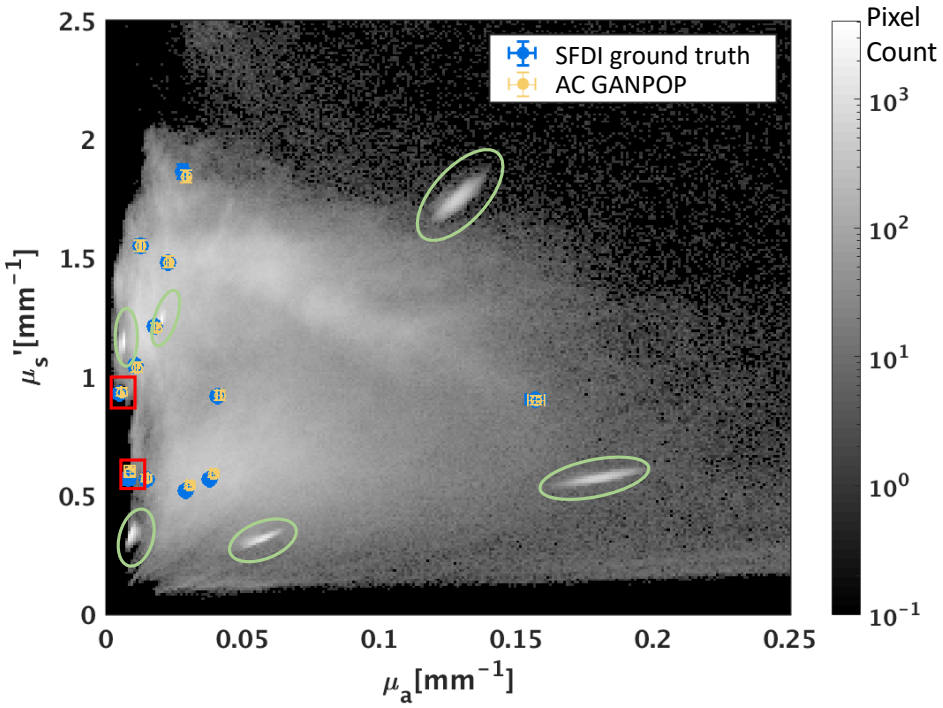

Figure 2-5. Scatter plot showing optical property pairs estimated by AC GANPOP compared to ground truth from conventional SFDI on 12 tissue phantoms. The 2D histogram in the background illustrates the distribution of training pixels among all optical property pairs, determined by SFDI. Green ellipses indicate dense pixel counts due to homogeneous phantoms used in training. Testing samples in the red box fell outside of the training range but were accurately predicted by AC GANPOP.

2.5.3 GANPOP Test on *Ex Vivo* Human Esophagus

GANPOP and SSOP were tested on the *ex vivo* human esophagus samples. NMAE scores were calculated for the two testing samples from each of four-fold cross validation iterations, and the average values from the four networks tested on a total of eight samples are reported in Fig. 2-7. Results from N_3, N_4, and SSOP are also compared to profilometry-corrected ground truth and shown in the same bar chart. On average, AC GANPOP produced approximately 60% higher accuracy than SSOP. Example optical property maps of a testing sample generated by N_1 are shown in Fig. 2-11(a).

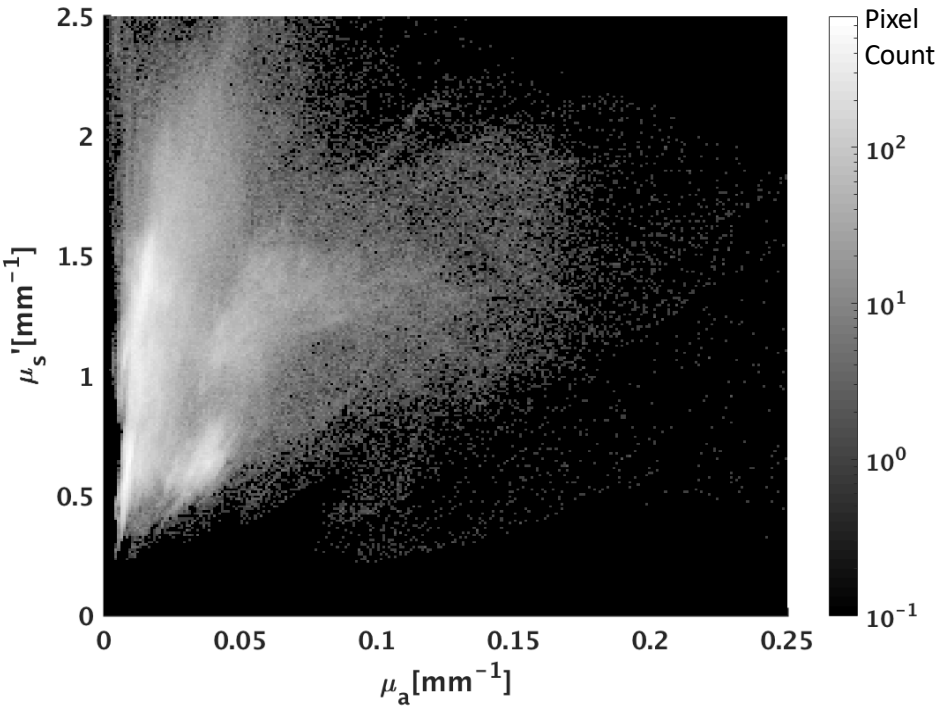

Figure 2-6. Histogram of optical property distribution of testing pixels from pig samples. Compared to training samples, pig tissues tested in this study had, on average, lower absorption coefficients and higher scattering coefficients.

2.5.4 GANPOP Test on Pig Samples

Each of the four GANPOP networks were tested on *ex vivo* esophagus and stomach samples from four pigs. Average NMAE scores for GANPOP and SSOP method were calculated for all eight pig tissue specimens (four esophagi and four stomachs) and are summarized in Fig. 2-7. Background regions, which were absorbing paper, were manually masked in the calculation, and the reported scores are the average values of 779,101 tissue pixels. Despite the fact that some testing samples had optical properties not covered by the training set (Fig. 2-6), AC GANPOP outperforms SSOP in terms of average accuracy and image quality (Fig. 2-11).

The networks were additionally tested on an *in vivo* pig colon. Average NMAE

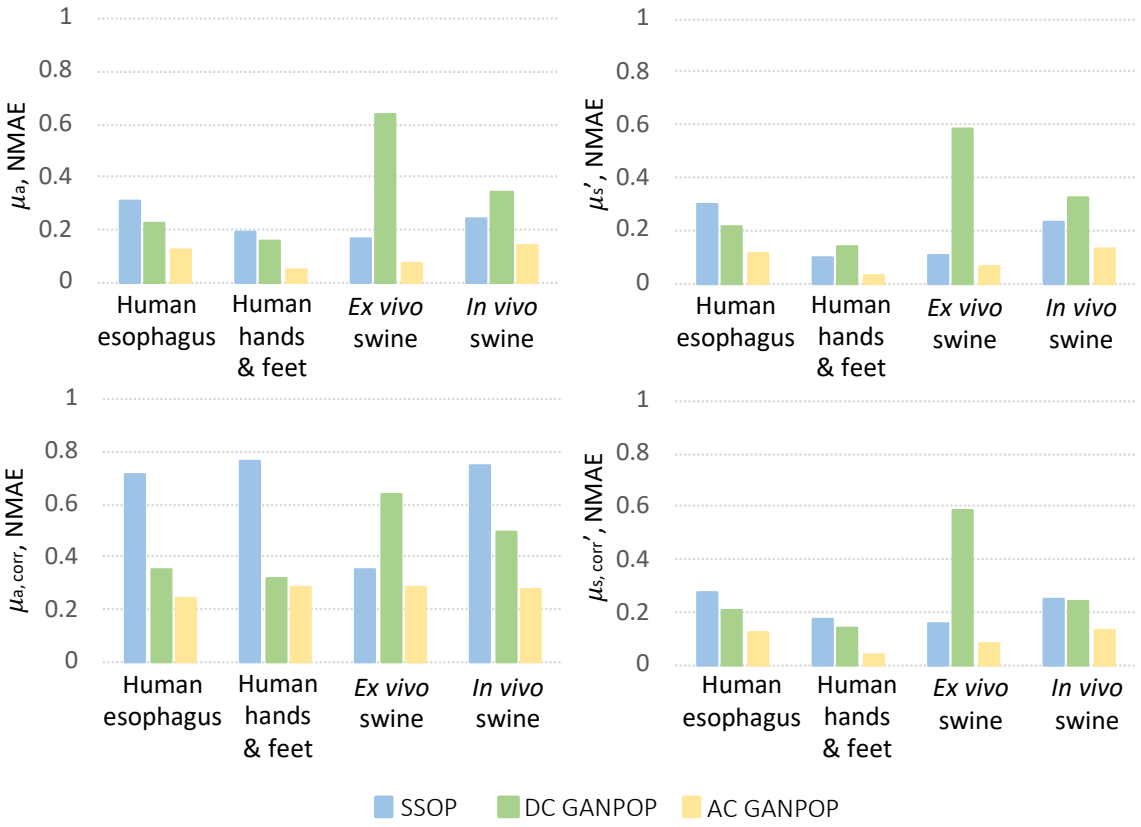

Figure 2-7. Accuracy of SSOP (blue), DC GANPOP (green) and AC GANPOP (yellow) in predicting optical properties of various types of samples. Average NMAE for absorption (left column) and scattering coefficients (right column) are reported. Top row shows the accuracy as compared to profile-uncorrected SFDI ground truth, and bottom row is against corrected ground truth. The same uncorrected SSOP data is used in both cases.

scores for GANPOP and SSOP are reported in Fig. 2-7 as average values of 118,594 pixels. The generated maps are shown in Fig. 2-11(c). AC GANPOP produces more accurate results than SSOP when compared to both uncorrected and profile-corrected ground truth data.

2.5.5 GANPOP Test on Hemisphere Phantoms

To prove that AC GANPOP has the potential to infer profilometry correction from a single image, N_3 was additionally tested on a hemisphere phantom with expected

Table 2-II. Performance comparison of SSOP, DC GANPOP, and AC GANPOP for full hand images, vessels, and background tissues.

	SSOP		DC GANPOP		AC GANPOP	
	μ_a NMAE	μ_s' NMAE	μ_a NMAE	μ_s' NMAE	μ_a NMAE	μ_s' NMAE
Vessels	0.1069	0.0600	0.1007	0.0968	0.0305	0.0206
Background	0.2311	0.1307	0.1345	0.1203	0.0407	0.0321
Hand overall	0.2262	0.1284	0.1339	0.1199	0.0404	0.0320

$\mu_a = 0.013mm^{-1}$ and $\mu_s' = 1.5mm^{-1}$. Shown in Fig. 2-8, AC GANPOP produces smaller errors than uncorrected SFDI, especially for angles greater than 30 degrees. Moreover, AC GANPOP results follow a similar error profile to the corrected ground truth. With a more sophisticated profilometry correction scheme, this error can be further minimized.

2.5.6 GANPOP Test on *In Vivo* Human Hands

To further demonstrate that GANPOP is able to accurately compute optical properties of inhomogeneous media, we segmented hand images into vessels and background tissue and calculated the respective NMAE (Table 2-II). Fig. 2-9 shows optical property maps of a representative human hand from SFDI, SSOP, and AC GANPOP, with vessels clearly visible in the absorption maps.

2.5.7 Decoupling Absorption and Scattering

To explore the capability of GANPOP to decouple the contributions of scattering and absorption to reflectance measurements, we compare SSOP, DC GANPOP, and AC

GANPOP to a baseline predictor that outputs the average scattering and absorption coefficients from training pixels with equal diffuse reflectance. For the human esophagus, examining all pixels with an $R_{d,0mm^{-1}} = 0.200 \pm 0.0025$, we find that the ground truth SFDI optical property measurements show a standard deviation of 26.7% for absorption and 25.0% for reduced scattering. The baseline change in optical properties between $R_{d,0mm^{-1}} = 0.200$ and 0.205 is much smaller–approximately 3.7% in absorption and 1.2% in reduced scattering. The normalized errors show that absorption and scattering errors are correlated for all methods (Fig. 2-10(e)-(h)). To assess the overall ability of each method to decouple optical properties, Fig. 2-10(i) shows the average optical property deviation of all optical property pairs for each reflectance value in the human esophagus testing samples. The SSOP error decreases with larger R_d, likely due to increased signal-to-noise ratios resulting from more detected photons. The baseline exhibits an improvement in accuracy that is generally correlated with the number of training pixels. DC GANPOP outperforms the baseline error at almost all values, indicating that its wide distribution of optical property estimates is consistently better than a constant pair prediction, and that it is effectively incorporating information about image content into each pixel prediction. AC GANPOP achieves the lowest error for all reflectances with more than 20 training pixels.

2.5.8 Comparative Analysis of Existing Deep Networks

Several deep learning architectures were explored for the purpose of optical property mapping, including conventional U-Net [44] and ResNet [77], both stand-alone and

integrated in a cGAN framework [65, 78]. The NMAE performance of each architecture was compared to AC GANPOP. All the networks were four-fold cross-validated, and the testing dataset included eight *ex vivo* human esophagi, four *ex vivo* pig GI samples, one *in vivo* pig colon, and eight *in vivo* hands and feet (Table 2-III).

2.6 Discussion

In this study, we have described a GAN-based technique for end-to-end optical property mapping from single structured (AC) and flat-field (DC) illumination images. Compared to the original pix2pix paradigm [65], the generator of our model adopted a fusion of U-Net and ResNet architectures for several reasons. First, a fully residual network effectively resolved the issue of vanishing gradients, allowing us to stably train a relatively deep neural network [78]. Second, the use of both long and short skip connections enables the network to learn from the structure of the images while preserving both low and high frequency details. The information flow both within and between levels is important for the prediction of optical properties, as demonstrated by the improved performance over a U-Net or ResNet approach. To further demonstrate the importance of skip connections in the network, we conducted an experiment with all skip connections removed and the network failed to converge. We also varied the number of convolutional layers by adding two additional layers in both the generator and the discriminator. Without skip connections, the deeper network did not converge. With skip connections, the network produced similar accuracy to the original GANPOP architecture. However, adding more layers means a higher computational cost and a

Table 2-III. Performance comparison of the proposed framework against model-based SSOP and other deep learning architectures when tested on profile-uncorrected data (N_1). Performance is measured in terms of normalized mean absolute error (NMAE).

Data type	SSOP		ResNet		UNet		ResNet-UNet		ResNet GAN		UNet GAN		**Proposed** (ResNet-UNet GAN)	
	μ_a	μ_s'	μ_a	μ_s'	μ_a	μ_s'	μ_a	μ_s'	μ_a	μ_s'	μ_a	μ_s'	μ_a	μ_s'
Human esophagus	0.312	0.298	0.192	0.136	0.144	0.136	0.185	0.129	0.201	0.140	0.148	0.143	**0.124**	**0.121**
***In vivo* pig colon**	0.171	0.112	2.032	0.145	0.251	0.186	1.533	0.145	1.953	0.133	0.190	0.152	**0.074**	**0.067**
***Ex vivo* pig GI tissue**	0.246	0.235	0.516	0.415	0.208	0.187	0.392	0.337	0.511	0.564	0.187	0.171	**0.143**	**0.133**
***In vivo* hands and feet**	0.194	0.101	0.337	0.070	0.100	0.066	0.250	0.068	0.643	0.162	0.089	0.056	**0.048**	**0.030**
Overall	0.231	0.187	0.769	0.192	0.176	0.144	0.590	0.170	0.827	0.250	0.154	0.131	**0.097**	**0.088**

longer time to train as it makes the model significantly larger (300 million parameters, as compared to 78 million in the original model).

As shown in Table 2-III, the inclusion of a discriminator significantly improved the performance of the fusion generator. This was especially apparent in the case for pig data, likely due to this testing tissue differing considerably from the training samples. We hypothesize that the cGAN architecture enforced the similarity between generated images and ground truth while preventing the generator from depending too much on the context of the image. Overall, the AC GANPOP method outperformed the other deep networks by a significant margin on all data types (Table 2-III). We additionally conducted ablation studies by isolating the $\mathcal{L}_1$ and $\mathcal{L}_2$ loss in Eq. 2.3. With only $\mathcal{L}_1$, the network became a standalone fusion generator, which performed poorly compared to using the proposed objective (ResNet-UNet versus ResNet-UNet GAN in Table 2-III). With only $\mathcal{L}_2$ loss, the network failed to converge. This is because $\mathcal{L}_1$ loss guides the training of the network after initialization, when there are large differences between predictions and ground truth. As the predictions become more accurate, $\mathcal{L}_2$ contributes more to the overall loss function. Additionally, we empirically found that a least squares GAN outperformed a conventional GAN when trained for 200 epochs. However, as discussed in [84], this improvement could potentially be matched by a conventional GAN with more training.

The training set used in this study is relatively small, including 6 human esophagi, 6 hands and feet, and 6 homogeneous phantoms. Effective training is achieved in this small dataset through several strategies. First, the incorporation of spectral

normalization stabilizes training and prevents mode collapse [79]. Second, we utilized patch-wise training and applied random stride sizes when extracting patches. This served to augment the dataset and provided random jitter. Moreover, training samples had heterogeneous optical properties, covering a wide range for both absorption and scattering (Fig. 2-5). We believe that this also helped with data efficiency as every pixel was used in learning the transformation from reflectance images to optical properties.

Compared to phantom ground truth in Fig. 2-5, AC GANPOP estimated optical properties with standard deviations on the same order of magnitude as conventional SFDI. Additionally, the AC GANPOP networks exhibited potential to extrapolate phantom optical properties that were not present in the training samples (highlighted by the red boxes in Fig. 2-5). This provides evidence that these networks have successfully learned the relationship between diffuse reflectance and optical properties, and are able to infer beyond the range of training data.

Fig. 2-7 shows that AC GANPOP consistently outperformed SSOP when tested on these types of data. From Fig. 2-6, it is evident that optical properties of the pig samples differed considerably from those of human esophagi used for training. Nevertheless, AC GANPOP exhibited more accurate estimation than the model-based SSOP benchmark. Moreover, a single network was trained for estimating both μ_a and μ_s' due to its lower computational cost and potential benefits in learning the relationships between the two parameters in tissues.

Table 2-II displays the accuracy of AC GANPOP compared to SSOP on different

regions a representative hand image, including vessels and background tissues. The NMAE values of these subtypes have a similar trend to those of the full images. Combined with the qualitative results shown in Fig. 2-9, this indicates that GANPOP is capable of not only accurately inferring optical properties of relatively homogeneous media, but also capturing subtle changes caused by transitions in tissue types.

Compared to SSOP, AC GANPOP optical property maps contain fewer artifacts caused by frequency filtering (Fig. 2-9 and 2-11). For both GANPOP and SSOP optical property estimation, a relatively large error is present on the edge of the sample. This is caused by the transition between tissue and the background, which poses problems for SFDI ground truth, and would be less significant for *in vivo* imaging. Artifacts caused by patched input are visible in GANPOP images, which can be reduced by using a larger patch size. However, this was not implemented in our study due to the size and the number of the specimens available for training. In our benchmarking with SSOP, we implemented the first version of the technique, which does not correct for sample height and surface angle variations. This allowed comparing identical input images for both SSOP and AC GANPOP.

In addition to training GANPOP models to estimate optical properties from objects assumed to be flat (N_1 and N_2), we trained networks that directly estimate profilometry-corrected optical properties (N_3 and N_4). For the same AC input, these models generated improved results over SSOP when tested on human and pig data. When compared against profile-corrected ground truth, they produced 47.3% less error for μ_a and 29.1% for μ_s' than did uncorrected output from N_1. Combined with

results shown in Fig. 2-8, this demonstrates that AC GANPOP is capable of inferring surface profile from a single fringe image and adjusting measured diffuse reflectance accordingly.

In experiment N_2 and N_4, when trained on DC illumination images, the GANPOP model became less accurate. It is important to note that any model that only considers DC reflectance values from pixels individually would be inherently limited in predicting the correct optical property pair among the infinite possibilities that would give the same reflectance measurement. However, since GANPOP is a content-aware framework that incorporates information from surrounding pixels in its prediction, it is possible to estimate optical properties from a more representative distribution than an approach that considers pixels in isolation. This hypothesis is supported by Fig. 2-10, which shows that DC GANPOP produces a wide range of estimates for a single diffuse reflectance and a lower average error than a baseline approach that gives a single optical property pair that minimizes errors for each reflectance in a training set. Therefore, we hypothesize that, when sufficiently trained on a certain tissue type, GANPOP has the potential to enable fairly rapid and accurate wide-field measurements of optical properties from conventional camera systems. This could be useful for applications such as endoscopic imaging of the GI tract, where the range of tissue optical properties is limited and modification of the hardware system is challenging.

The generator models trained in this study were 600MB in size with 78 million parameters. Each iteration was trained on approximately 500 patches, and this

process took 3 hours on an NVIDIA Tesla P100 GPU. In terms of speed, GANPOP requires capturing one sample image instead of six, thus significantly shortening data acquisition time while avoiding image artifacts due to motion or change in ambient light. For optical property extraction, the model developed here without optimization takes approximately 0.04 s to process a 256 $\times$ 256 image on an NVIDIA Tesla a P100 GPU. Therefore, this technique has the potential to be applied in real time for fast and accurate optical property mapping. In terms of adaptability, random cropping ensures that our trained models work on any 256 $\times$ 256 patches within the field of view. Additionally, while the models were trained on the same calibration phantom at 660 nm, they could theoretically be applied to other references or wavelengths by scaling the average $M_{DC,ref}$ and $M_{AC,ref}$.

For future work, a more generalizable model that would work on a range of imaging systems could be trained using domain adaptation techniques. A wider range of optical properties could also be incorporated into the training set, though this would inevitably incur a higher computational cost and necessitate a much larger dataset for training. Another future direction is to explore the applications of GANPOP in situations where SFDI data is difficult to acquire. Findings from this study indicate that training on a relatively small set of images can enable a GANPOP generator that is accurate and robust. Thus, in cases where acquiring training data is expensive or laborious, we hypothesize that a small dataset containing relatively few samples would be sufficient for enabling accurate predictions. Moreover, although all input images used here were acquired at an approximately-constant working distance, GANPOP could be modified

to work for a variety of imaging geometries. Incorporating monocular depth estimates into the prediction may enable GANPOP to account for large differences in working distance [85, 86]. This could be particularly useful for endoscopic screening where constant imaging geometries are difficult to achieve. Having a model trained on images at multiple wavelengths, this technique can be modified to provide critical information in real time, such as tissue oxygenation and metabolism biomarkers. Accuracy in this application may also benefit from training adversarial networks to directly estimate these biomarkers rather than using optical properties as intermediate representations. By similar extension, future research may develop networks to directly estimate disease diagnosis and localization from structured light images.

2.7 Conclusion

We have proposed a deep learning-based approach to optical property mapping (GANPOP) from single snapshot wide-field images. This model utilizes a conditional Generative Adversarial Network consisting of a generator and a discriminator that are iteratively trained in concert with one another. Using SFDI-determined optical properties as ground truth, AC GANPOP produces significantly more accurate optical property maps than a model-based SSOP benchmark. Moreover, we have demonstrated that DC GANPOP can estimate optical properties with conventional flat-field illumination, potentially enabling optical property mapping in endoscopy without modifications for structured illumination. This method lays the foundation for future work in incorporating real-time, high-fidelity optical property mapping and

quantitative biomarker imaging into endoscopy and image-guided surgery applications.

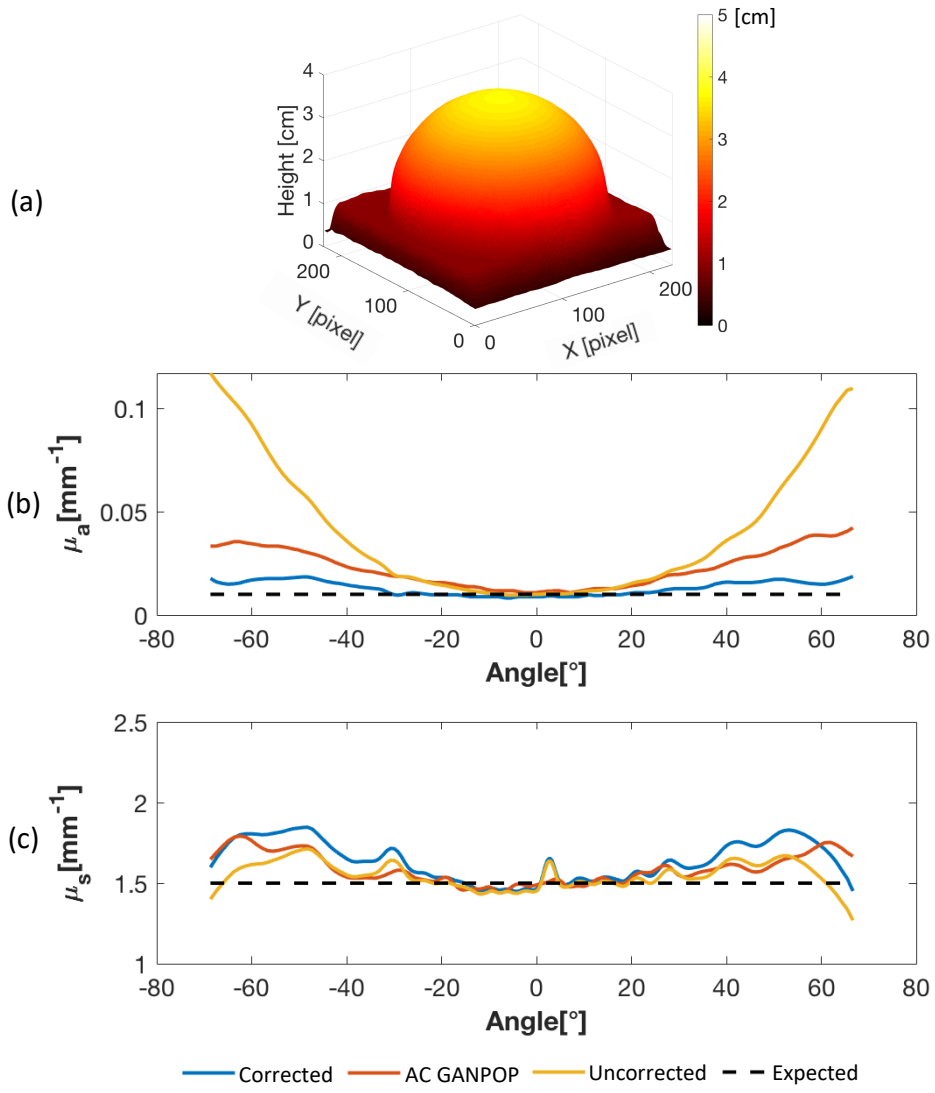

Figure 2-8. (a): Height map of a hemisphere phantom measured by SFDI profilometry; (b) and (c): Absorption and scattering coefficients measured by AC GANPOP (N_3), profile-corrected (ground truth) and uncorrected SFDI compared to expected values.

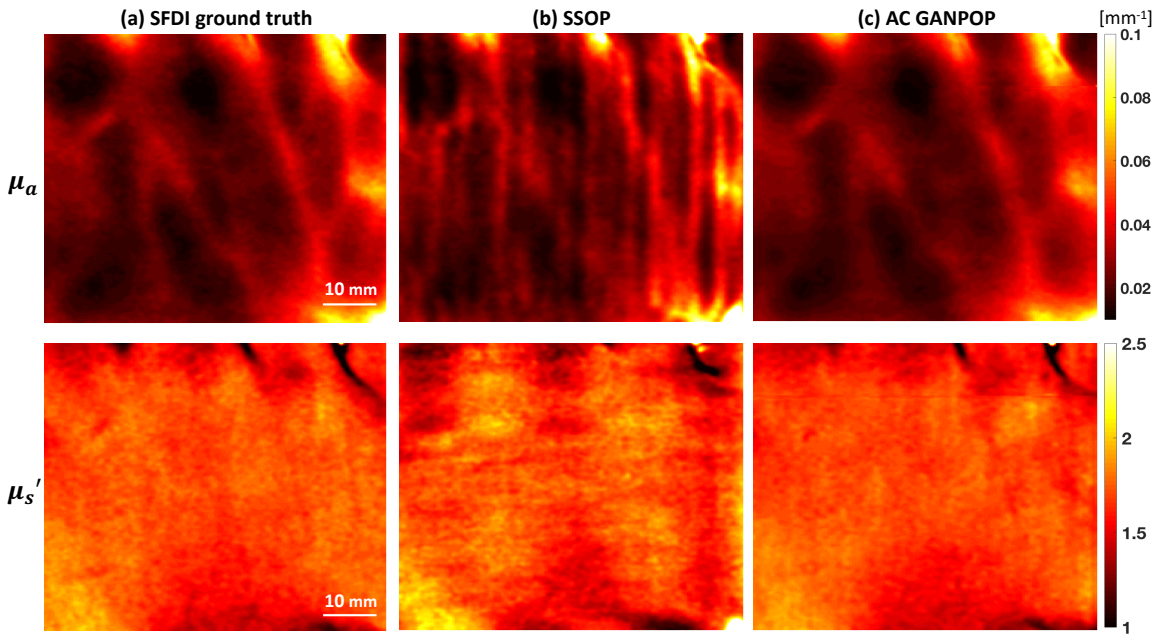

Figure 2-9. Absorption (top) and reduced scattering (bottom) maps of the back of a hand generated by (a) SFDI ground truth, (b) SSOP, and (c) AC GANPOP. GANPOP result closely resembles the ground truth in terms of pixel accuracy and image quality, even for fine vascular structures.

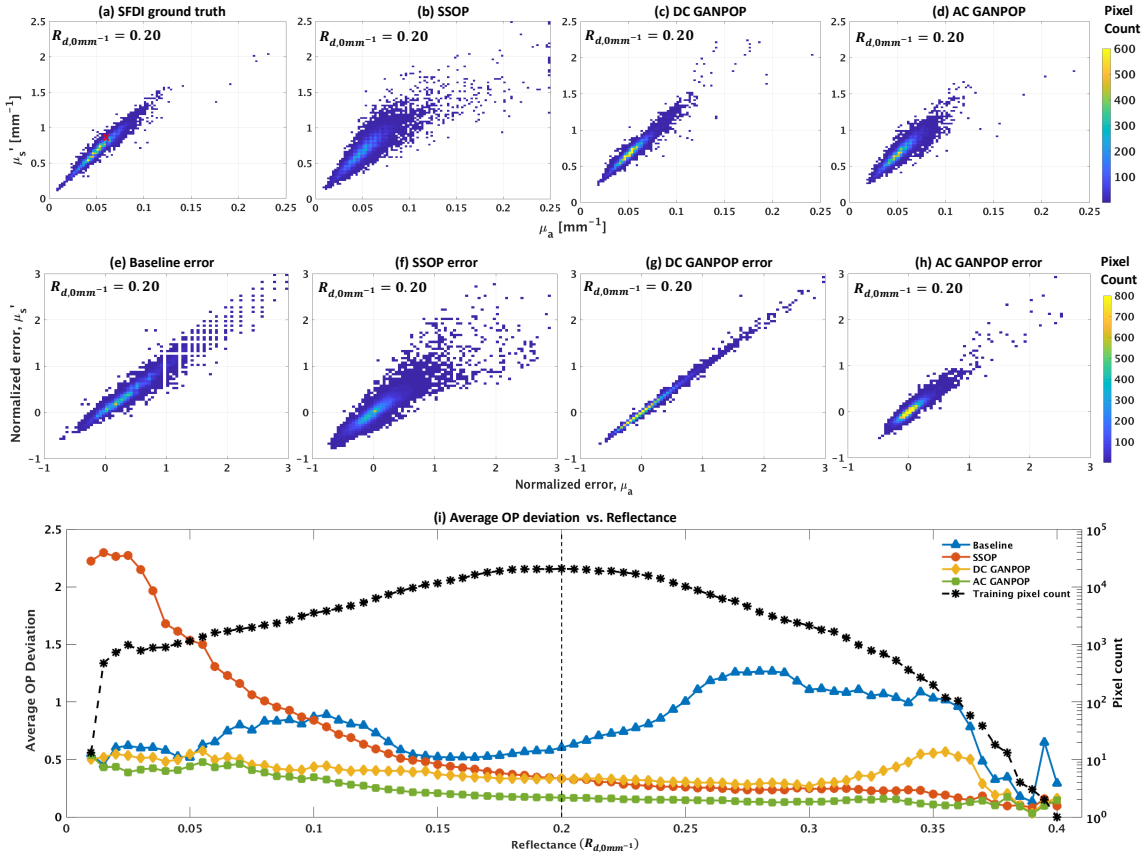

Figure 2-10. Top row: optical property distributions of esophagus test samples estimated by: (a) SFDI ground truth, (b) SSOP, (c) DC GANPOP, and (d) AC GANPOP, at a constant reflectance of $R_{d,0mm^{-1}} = 0.20$. Red cross in (a) marks the baseline measurement, which is the average value of all training optical property pairs. Middle row: corresponding error histograms from (e) baseline, (f) SSOP, (g) DC GANPOP, and (h) AC GANPOP. Bottom row: (i) plot of average average optical property (OP) deviation of all optical property pairs for a given DC reflectance. Each average OP deviation is calculated as the average distance of all test points from ground truth at the corresponding reflectance level. Dotted line in (i) indicates where (a)-(h) are sampled.

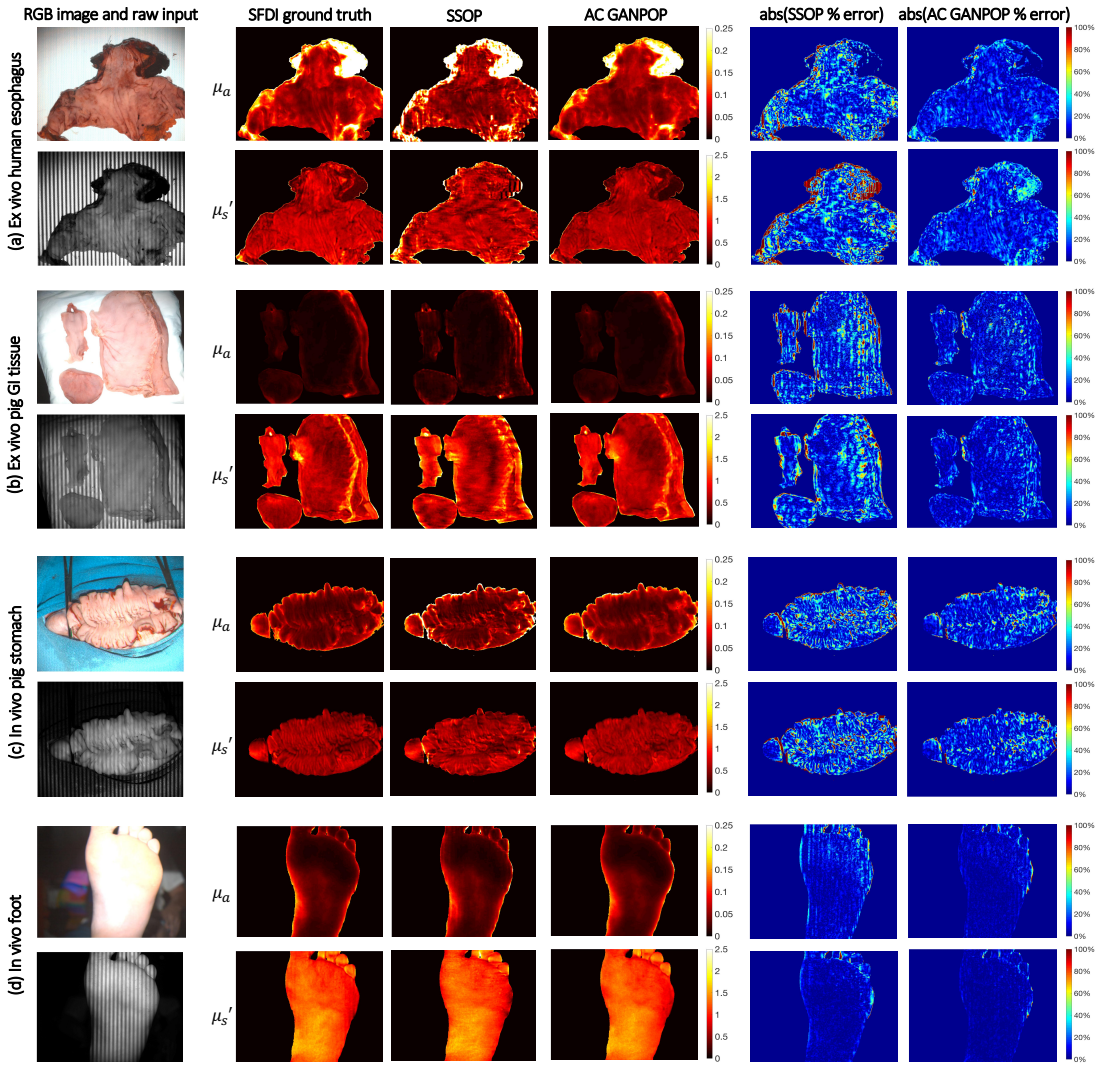

Figure 2-11. Example results for AC input to non-profilometry-corrected optical properties (N_1). From left to right: RGB image and raw structured illumination image, SFDI ground truth, SSOP output, AC GANPOP output, absolute percent error map between SSOP and ground truth, and absolute percent error map between AC GANPOP and ground truth. From top to bottom: (a) *ex vivo* human esophagus, (b) *ex vivo* pig stomach and esophagus, (c) *in vivo* pig colon, and (d) *in vivo* human foot. Optical properties are measured in mm^{-1}.

Chapter 3

Rapid Tissue Oxygenation Mapping from Snapshot Structured-light Images with Adversarial Deep Learning

This chapter contains work originally published in [49, 51]:

West, 2021.

3.1 Introduction

In the previous chapter, a data-drive approach was presented for high-fidelity optical property mapping from single-shot SFDI images. A natural extension of this work would be to estimate chromophore concentrations using a similar approach. One important biomarker related to chromophore concentrations is tissue oxygenation (StO_2). Tissue oxygenation is a measure of the amount of oxygen in biological tissue, and is often estimated by computing the fraction of oxygenated hemoglobin over total hemoglobin. StO_2 is a useful clinical biomarker for tissue viability, the continuous monitoring of which is valuable for surgical guidance and patient management [87, 88]. Abnormal levels of StO_2 are indicative of many pathological conditions, such as sepsis, diabetes, and chronic obstructive pulmonary disease [89–91].

One of the most widely used techniques to measure physiological oxygen levels is pulse oximetry. Despite its ubiquity, robustness, and low cost, pulse oximetry requires a pulsatile arterial signal and only provides a systemic measure of oxygenation [92, 93]. The majority of existing devices for local assessment of StO_2 are based on near-infrared (NIR) spectroscopy. NIR spectroscopy quantifies the concentrations of oxygenated and de-oxygenated hemoglobin by characterizing tissue absorption of light at wavelengths typically between 650 and 1000 nm [94]. Similar to pulse oximetry, spectroscopic probes require direct contact with tissue. These measurements can be noisy as they are sensitive to pressure and sample movement [95–97]. Compared to

tissue probes, spectroscopic imaging techniques are advantageous as they provide non-contact readings of oxygen saturation at a high spatial resolution and large field of view [98]. Nevertheless, continuous-wave NIR spectroscopy assumes constant scattering, which could be a source of error as scattering coefficients are often spatially non-uniform. Therefore, to accurately determine oxygen saturation, it is imperative to separate the effect of optical properties, including absorption (μ_a) and reduced scattering coefficients (μ_s'). Spectrally constrained reconstructions have been shown to be useful in measuring chromophore concentrations and μ_s', but this technique tends to require complex instrumentation and suffer from limited fields of view [99].

In recent years, SFDI has emerged as a promising technique for measuring tissue optical properties. SFDI quantifies optical properties by projecting structured light and characterizing the modulation transfer function of tissues in the spatial frequency domain [2]. Oxygenation can subsequently be determined by fitting chromophore concentrations to the measured absorption coefficients using the Beer-Lambert law. In addition to isolating the effect of tissue scattering, SFDI is a wide-field, non-contact technique that can be implemented using a simple setup, which includes a camera and a projector. These advantages make SFDI suitable for many clinical applications that necessitate accurate StO_2 measurements, such as burn wound assessment [100, 101], pressure ulcer staging and risk stratification [102], image-guided surgery [93, 97], and cancer therapy evaluation [103].

Despite its growing use in various applications, there are several factors that limit the clinical translation of SFDI. First, compared to probe-based oximetry, SFDI

components are costly. For example, digital micromirror devices or spatial light modulators are often used to produce programmable structured illumination. Second, SFDI requires carefully-controlled imaging geometries, which can be difficult to achieve in clinical settings. Moreover, conventional SFDI requires six images per wavelength (0, $\frac{2}{3}\pi$, and $\frac{4}{3}\pi$ phase offsets at two spatial frequencies) and a pixel-wise lookup table (LUT) search to generate a single optical property map. For robust oxygenation estimates, absorption coefficients at a minimum of two wavelengths are needed, and an additional least square fitting step is performed (Fig. 3-1(a)) [104]. Previous work has shown that real-time imaging can be achieved with single-snapshot acquisition [105] and either an optimized LUT [106] or a machine learning inversion method [45]. However, single image acquisition and frequency filtering often result in image artifacts and high per-pixel error [107]. Therefore, wide-field, rapid, and accurate StO_2 measurement still remains a challenge.

In recent years, convolutional neural networks (CNNs) have emerged as a powerful tool in many medical imaging-related tasks [108, 109]. By employing convolutional filters followed by dimension reduction and rectification, CNNs are capable of extracting high-level features and interpreting spatial structures of input images [110]. For image translation tasks, generative adversarial networks (GANs) improve upon conventional CNNs by utilizing both a generator and a discriminator [111] to effectively model a complex loss function. The two components are trained simultaneously, with the generator learning to produce realistic output and the discriminator to classify the generator output as real or fake. Recently, GANs have been employed to predict

optical properties from single structured illumination images (GANPOP) [3]. As a content-aware network, this technique significantly improves upon the accuracy of model-based single snapshot techniques in estimating optical properties. However, to compute StO_2 with the GANPOP approach, multiple wavelength-specific networks must be run to first estimate absorption coefficients, followed by chromophore fitting, which compounds errors and increases computational demand (Fig. 3-1(b)). In this study, we present an end-to-end technique for computing StO_2 directly from structured-illumination images using generative adversarial networks (OxyGAN). OxyGAN maps StO_2 from single snapshot images from 659 and 851 nm sinusoidal illumination. We train generative networks to estimate both uncorrected and profile-corrected StO_2 and compare the performance of the end-to-end architecture versus intermediately calculating optical properties. We accelerate OxyGAN model inference by importing the framework into NVIDIA TensorRT for efficient deployment. Finally, we demonstrate real-time OxyGAN by recording its estimation over the course of a three-minute occlusion experiment.

3.2 Methods

For training and testing of OxyGAN, single structured illumination images were acquired at two different wavelengths (659 and 851 nm) and paired with registered oxygenation maps. Ground truth oxygenation is obtained using the absorption coefficients measured by conventional SFDI at four wavelengths (659, 691, 731, and 851 nm). Experiments were conducted using both profile-corrected [30, 34] and

uncorrected ground truth. The training set included human *ex vivo* esophagus samples and *in vivo* feet. OxyGAN was evaluated using unseen tissues of the same type as the training data (human *in vivo* feet) and in different tissue types (human *in vivo* hands and a pig *in vivo* colon). Its performance was additionally compared with single-snapshot optical properties (SSOP) [107, 112] as a model-based benchmark that utilizes a single structured-light image.

3.2.1 Ground Truth Tissue Oxygenation

In this study, conventional SFDI was used to obtain ground truth StO_2 maps. At each wavelength, structured illumination images were captured using a commercial SFDI system (Reflect RS, Modulim Inc.) at two spatial frequencies (0 and 0.2 mm^{-1}) and three phase offsets (0, $\frac{2}{3}\pi$, and $\frac{4}{3}\pi$). The process was implemented for both the sample of interest and a reference phantom. The acquired images were then demodulated and calibrated against the response of the reference phantom at each frequency. The DC (0 mm^{-1}) and AC (0.2 mm^{-1}) diffuse reflectance of the sample were fit to a LUT generated by White Monte Carlo simulations [28]. This pixel-wise LUT search resulted in an optical property map of the sample, which consisted of scattering corrected absorption (μ_a) and reduced scattering (μ_s') coefficients. In experiments where profile-corrected ground truth was used, we also implemented height and surface normal angle correction [30, 34]. With μ_a measured at four different wavelengths (659, 691, 731, and 851 nm), we subsequently estimated chromophore concentrations using

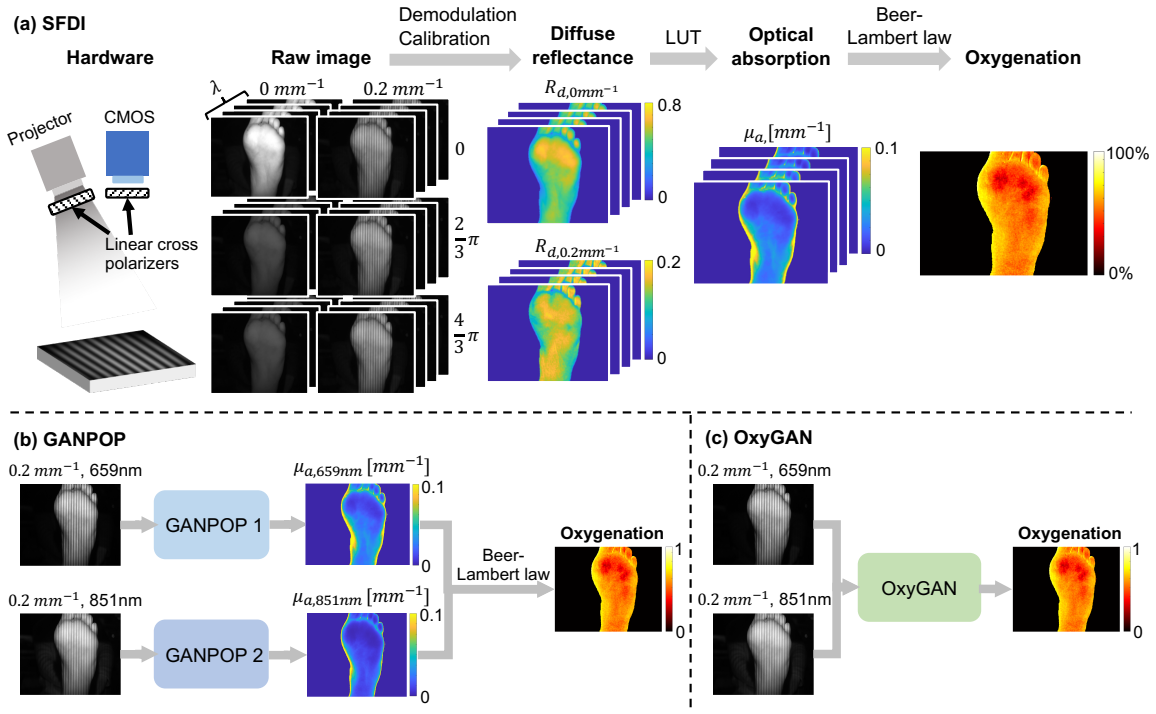

Figure 3-1. Comparison of (a) SFDI, (b) GANPOP, and (c) OxyGAN StO_2 techniques. Ground truth SFDI uses 24 input images (2 spatial frequencies, 3 phases, and 4 wavelengths), while GANPOP and OxyGAN use 2 input images (1 spatial frequency, 1 phase, 2 wavelengths). SFDI and GANPOP absorption maps that are subsequently fit to basis chromophores to estimate StO_2. OxyGAN directly calculates oxygen saturation with a single network, reducing compounding errors and processing time.

the Beer-Lambert Law (Eq. 1.7). Oxygen saturation was then calculated as:

$$stO_2 = \frac{c_{O_2Hb}}{c_{O_2Hb} + c_{HHb}}, \tag{3.1}$$

where c_{O_2Hb} and c_{HHb} represent the concentrations of oxygenated and de-oxygenated hemoglobin, respectively. We estimated ground truth oxygenation maps using absorption coefficients at all the near-infrared wavelengths available in the system (659, 691, 731, and 851 nm).

3.2.2 SSOP Benchmark

In this study, we implemented SSOP as a model-based benchmark. Briefly, this method calculates tissue optical properties from single structured-illumination images by 2-D filtering in the frequency domain [107, 112]. We applied anisotropic low pass filtering using a sine window and high pass filtering using a Blackman window [112]. The absorption coefficients measured by SSOP at 659nm and 851nm were substituted into Eqs. 1.7 and 3.1 to estimate StO_2.

3.2.3 OxyGAN Framework

In this study, we pose StO_2 mapping as an image-to-image translation task. OxyGAN uses an adversarial training framework to accomplish this task (Fig. 3-2). Specifically, OxyGAN is a conditional generative adversarial network (cGAN) that consists of two convolutional neural networks–a generator and a discriminator. Both networks are conditioned on the same input data, which are single structured-light images in our case. First proposed in [64], the cGAN structure has been shown to be an effective solution to a wide range of image-to-image translation problems [113]. While conventional single-network CNNs require simple, hand-crafted loss functions, cGANs can be more generalizable because the discriminator can effectively learn a complex loss function.

For the OxyGAN generator, we implement a novel fusion network that combines the properties of U-Net and ResNet (Fig. 3-2) [44, 77]. Similar to a U-Net, the OxyGAN generator is an encoder-decoder setup with long skip connections between the two

branches on the same level. However, OxyGAN also includes short skip connections within each level and replaces U-Net concatenation with additions, making the network fully residual [114, 3]. The residual blocks on each level consist of five 3×3 convolutional layers, with residual additions between the outputs of the first and the fourth layer. Dimension reduction on the encoder side and expansion on the decoder side is achieved with 2×2 maxpooling and 3×3 up-convolutions, respectively. We use regular ReLUs for the encoder and leaky ReLUs with a slope of 0.2 for the decoder. A final 3×3 convolution followed by a $Tanh$ activation function is applied to generate the output. The discriminator is a three-layer PatchGAN with leaky ReLUs (slope = 0.2) [113], which results in a receptive field of 70×70 pixels. To stabilize the training process, we incorporated spectral normalization after each convolution layer in both the generator and the discriminator [115]. We use use an adversarial loss of:

$$\mathcal{L}_{\text{cGAN}}(G, D) = \mathbb{E}_{x,y}[\log(D(x, y))] + \mathbb{E}_{x}[\log(1 - D(x, G(x)))], \tag{3.2}$$

where G is the generator ($G : X \rightarrow Y$) and D is the discriminator [113]. During training, G tries to minimize this objective while its adversary, D, tries to maximize it. The discriminator is trained to determine if a given pair of images forms a correct reconstruction for a given input. This classification is made from data that includes the current input-ground truth pair and an image pair randomly sampled from a buffer of 64 previously-generated pairs. Additionally, an L_1 loss is included to improve the generator performance and training stability (Eq. 2.2). The full objective function is the same as Eq. 2.3, where λ is a hyperparameter that controls the weight of the

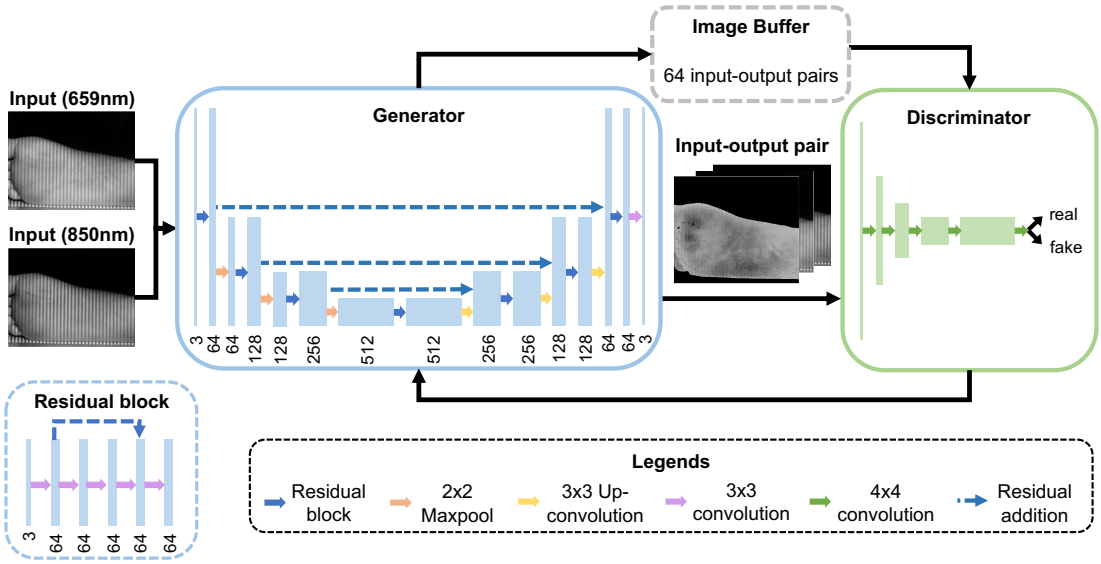

Figure 3-2. The OxyGAN framework. OxyGAN produces StO_2 maps directly from single-phase SFDI images with 659nm and 851nm illumination. The generator is a fusion network that combines the properties of U-Net and ResNet. The number under each block describes the number of channels. The discriminator is a PatchGAN classifier with a receptive field of 70×70 pixels. The discriminator trains to classify the current image pair versus an input-output pair randomly sampled from a pool of 64 previously generated images.

L_1 loss term and was set to 60. OxyGAN models solved this objective using an *Adam* solver with a batch size of 1 [81]. G and D weights were both initialized from a Gaussian distribution with a mean and standard deviation of 0 and 0.02, respectively. These models were trained for 200 epochs, and a constant learning rate of 0.0002 was used for the first 100 epochs. The learning rate was linearly decreased to 0 for the second half of the training process. The full algorithm was implemented using Pytorch 1.0 on Ubuntu 16.04 with a single NVIDIA Tesla P100 GPU on Google Cloud [3].

3.2.4 Data Split and Augmentation

In this study, we conducted separate experiments to estimate both uncorrected (StO_2) and profile-corrected oxygenation ($StO_{2,corr}$) from the same single-snapshot structured light image input. These networks were trained and tested on 256×256-pixel patches paired with registered oxygenation maps. To generate training datasets, each 520×696 image was segmented at a random stride size, which resulted in approximately 30 image pairs per sample. The input data was arranged in a way so that it efficiently utilized the three image channels normally used for color (Fig. 3-3). The first and second channel are flat-field corrected, single-phase illumination images at 659 and 851nm, respectively. To account for system drift over time, we included the ratio between demodulated AC (M_{AC}) and DC magnitude (M_{DC}) of the reference phantom in the third channel. Reference measurements were taken the same day as the tissue measurements, in the same way as conventional SFDI workflows. As shown in Fig. 3-3, the ratios at 659 and 851nm alternate in a checkerboard pattern to account for any spatial variations.

To prevent overfitting of the models, we augmented the training data by flipping the images horizontally or vertically. During each epoch, both flipping operations occurred with a 50% chance and were independent of each other. Data augmentation was important for this study because of the small size of the training set, and because the testing set includes new object types never seen in training. Additionally, since the classification task of the discriminator was easier than the generator, we applied the one-sided label smoothing technique when training the discriminator. In short,

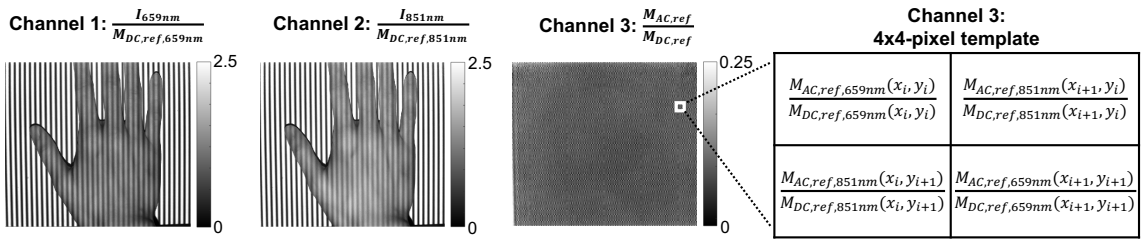

Figure 3-3. Images of the three input channels. Channel 1 and 2 are flat-field corrected single-snapshot SFDI images at 659nm and 851nm. Channel 3 gives ratios between demodulated AC ($M_{AC,ref}$) and DC magnitude ($M_{DC,ref}$) of the reference phantom. The ratios at 659nm and 851nm are alternated to form a checkerboard pattern, as shown in the 4×4-pixel template on the right.

the positive (real) targets with a value of 1 were replaced with a smoothed value (0.9 in our case). This was implemented to prevent the discriminator from becoming overconfident and using only a small set of features when classifying output [116].

3.2.5 Samples

The training set of OxyGAN models included eight *ex vivo* human esophagectomy samples [82] and four *in vivo* human feet, which resulted in approximately 1200 image pairs after augmentation. The testing set consisted of two *in vivo* human hands and feet and an *in vivo* pig colon. All models were cross-validated by training on four of the six feet and testing on the remaining two each time. All summary results reported indicate the average performance of these 3 sets of trained networks. OxyGAN models never see data from hands or *in vivo* pig colon in training.

We additionally recorded a 400 second video of a healthy volunteer's hand during an occlusion study. We first applied a household pressure cuff (Walgreens Manual Inflate Blood Pressure Kit) to the upper arm of the volunteer and imaged the hand at

baseline for a minute. Then, the cuff pressure was increased to 200mmHg to occlude the arm for approximately 3 minutes. The pressure was then released and the hand was imaged for another 2.5 minutes. Single-phase sinusoidal illumination was used, which alternated between 659nm and 851nm so that oxygenation could be measured at each time point ($\Delta t = 0.73s$). To obtain ground truth oxygenation, conventional six-image SFDI was implemented every 25 seconds, resulting in 15 measurements in total.

In this study, the protocols for *in vivo* imaging of human hands and feet (IRB00214149) and *ex vivo* imaging of esophagetomy samples (IRB00144178) were approved by Johns Hopkins Institutional Review Board. The *in vivo* imaging of the pig colon (SW18A164) was approved by Johns Hopkins Animal Care and Use Committee.

3.2.6 Performance Evaluation

In this study, we benchmarked OxyGAN by comparing it to SSOP. We additionally compared OxyGAN to an approach using GANPOP networks to first predict optical properties at 659nm and 851nm and subsequently fitting the concentrations of oxygenated and de-oxygenated hemoglobin using the Beer-Lambert law. These GANPOP networks were trained on the same dataset as OxyGAN with cross validation. The performance of all three methods was evaluated using Normalized Mean Absolute Error (NMAE), which is equivalent to absolute percentage error:

$$NMAE = \frac{\sum_{i=1}^{N} |StO2_i - StO2_{i,GT}|}{\sum_{i=1}^{N} StO2_{i,GT}}. \tag{3.3}$$

$StO2_i$ and $StO2_{i,GT}$ are predicted and SFDI ground-truth oxygen saturation, respectively. N is the total number of pixels. All testing datasets were manually masked to only include pixels that sampled the object.

3.3 Results

The average NMAEs are reported in Table 3-I for SSOP, U-Net, GANPOP, and OxyGAN tested on human feet, hands, and *in vivo* pig colon. The hands and feet are from different healthy volunteers with a wide range of pigmentation levels (Fitzpatrick skin types 1-5). It is worth emphasizing that the *in vivo* hands and pig colon were completely new tissue types that were not represented in the training set. On average, OxyGAN outperforms SSOP and GANPOP in accuracy by 24.04% and 6.88%, respectively, compared to uncorrected SFDI ground truth. Compared to profile-corrected ground truth, the improvement of OxyGAN over SSOP and GANPOP becomes 24.89% and 24.76%, respectively. The table also reports the amount of time it takes for each method to generate a 512×512 StO_2 map during testing. OxyGAN inference is approximately 2 times faster than a U-Net and 8 times faster than the GANPOP approach.

Figure 3-4 compares the results of profile-corrected SFDI, SSOP, U-Net, and OxyGAN applied to a sample of each testing tissue type. Lower errors and fewer image artifacts are observed in OxyGAN results. Error plots highlight the fringe artifacts commonly observed parallel to the illumination patterns in SSOP. Interestingly, U-Net collapses when applied to the human hand, which is a tissue type not included in the

Table 3-I. NMAE of StO_2 predicted by SSOP, U-Net, GANPOP, and OxyGAN compared to both uncorrected and profile-corrected SFDI ground truth. Runtime is the amount of time each method takes to generate a 512×512 oxygenation map

	vs. uncorrected SFDI				vs. profile-corrected SFDI				Runtime
	Feet	Hands	Pig	*Overall*	Feet	Hands	Pig	*Overall*	
SSOP	0.0396	0.0430	0.1508	*0.0778*	0.0601	0.0672	0.1404	*0.0892*	-
U-Net	0.1399	0.1762	0.1117	*0.1426*	0.1676	0.3387	0.1430	*0.2164*	0.08 s
GANPOP	0.0438	0.0381	0.1085	*0.0635*	0.0795	0.0655	0.1220	*0.0890*	0.32 s
OxyGAN	**0.0358**	**0.0335**	**0.1080**	***0.0591***	**0.0536**	**0.0466**	**0.1007**	***0.0670***	**0.04 s**

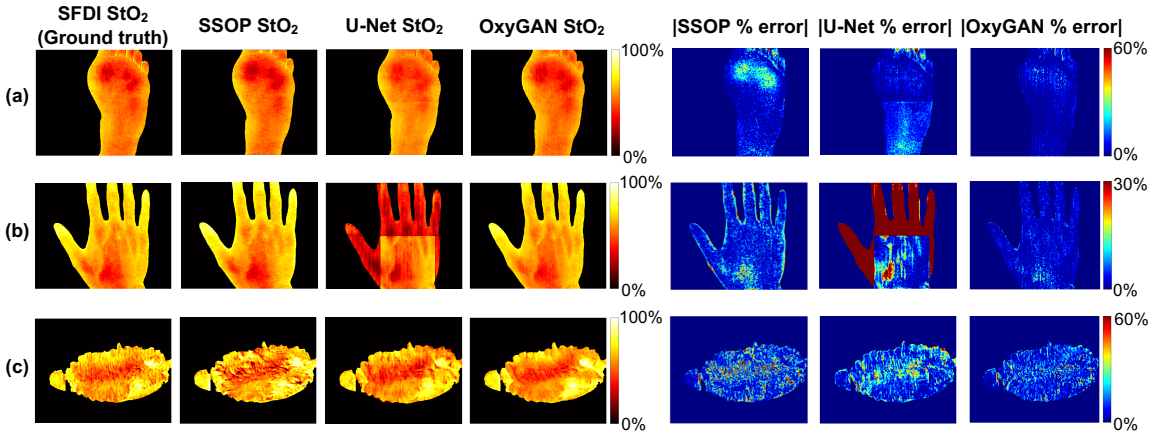

Figure 3-4. Comparison of profile-corrected SFDI, SSOP, U-Net, and OxyGAN StO_2 results. (a) *in vivo* human foot; (b) *in vivo* human hand; (c) *in vivo* pig colon.

training set, while OxyGAN still produces accurate results. This shows that OxyGAN is a more generalizable approach that is suitable for small datasets. In addition, all three techniques exhibited higher errors in the pig colon, which had more complex surface topography and made single-snapshot predictions more difficult.

We additionally implemented OxyGAN on a video of a volunteer's stationary hand during an occlusion study (Fig. 3-5). The average oxygen saturation was calculated for a region of interest highlighted by the red box in Fig. 3-5(c) and compared to the SFDI ground truth in Fig. 3-5(d). OxyGAN accurately measures a large range of oxygenation values and shows strong and stable agreement with conventional SFDI.

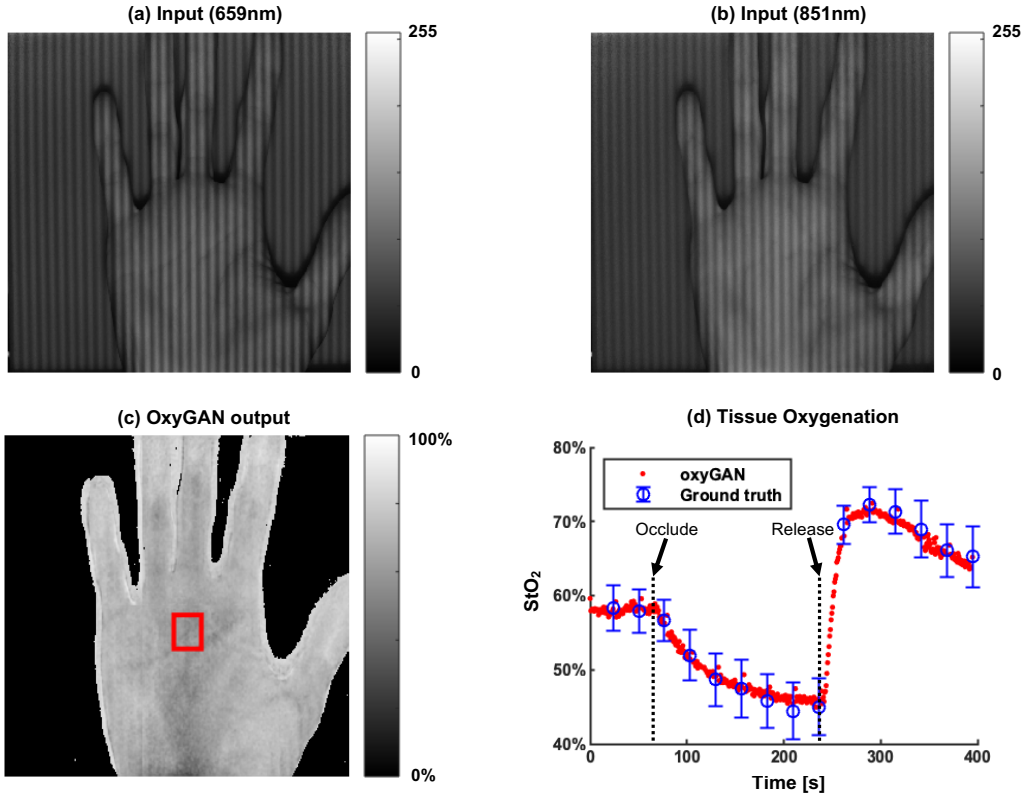

Figure 3-5. Video of the occlusion-release experiment. (a) Input at 659nm; (b) Input at 851nm; (c) StO_2 map predicted by OxyGAN; (d) StO_2 trend measured by OxyGAN and ground truth SFDI over time.

Finally, we implemented OxyGAN on a video of a moving hand (Fig. 3-6). OxyGAN demonstrates robustness to motion, producing accurate results with standard deviations of the same magnitude as ground truth SFDI. Importantly, the StO_2 measurement speed and artifacts are limited by the image acquisition speed rather than the OxyGAN speed in these cases.

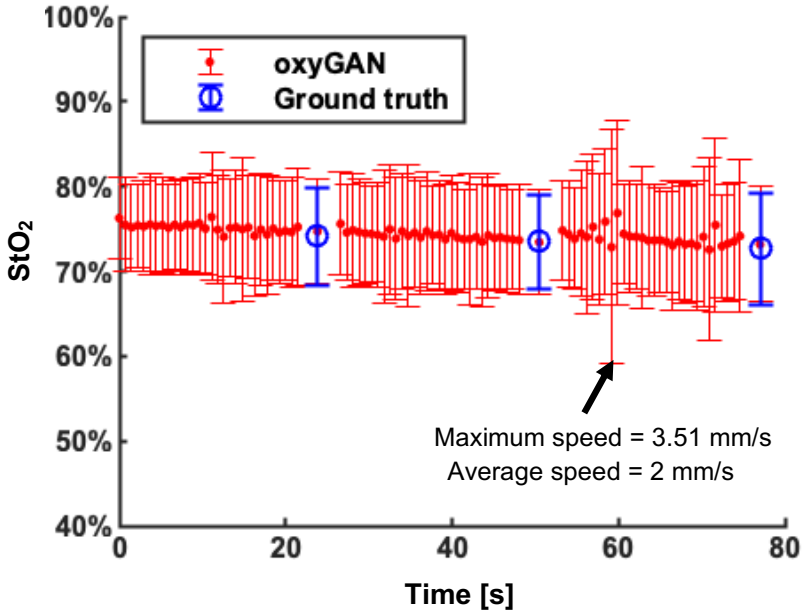

Figure 3-6. Trends of StO_2 predicted by OxyGAN and ground truth SFDI of a moving hand. Errors arise from movement of the object between sequential acquisition of structured light images at the two wavelengths and could be reduced with dual-color imaging.

3.4 Discussion

In this study, we have described a fast and accurate technique for estimating wide-field tissue oxygenation from single-snapshot structured illumination images using generative adversarial networks. As shown in Table 3-I, OxyGAN accurately measures oxygenation not only for sample types represented in the training set (human feet), but also for unseen sample types (human hands and pig colon). This supports the possibility that OxyGAN can be robust and generalizable. The occlusion video (Fig. 3-5) and the moving hand video (Fig. 3-6) further demonstrate the ability of OxyGAN to accurately measure a wide range of tissue oxygenation levels and detect changes over time.

Compared to training separate GANPOP networks to first estimate absorption coefficients, OxyGAN produces an average improved accuracy of 15.8%. Moreover, a greater improvement is observed in profile-corrected experiments. One potential explanation for this is that the errors in absorption coefficients due to uncertainties in profilometry estimation propagate and result in a larger error in oxygenation measurements. Additionally, compared to separate GANPOP models, the end-to-end OxyGAN approach requires only one network and bypasses the Beer-Lambert fitting step, thus greatly reducing the computational cost for training and inference. For example, training a network on 350 patches took approximately 2.2 hours, or 40 seconds per epoch on an NVIDIA Tesla P100 GPU. Training separate GANPOP networks would take double the amount of time and memory. To achieve real-time StO_2 mapping, we first converted the trained model to Open Neural Network Exchange (ONNX) format. We then imported the ONNX model into NVIDIA TensorRT 7 for reduced latency and optimized inference. For testing, OxyGAN inference on a Tesla P100 takes approximately 0.04s to generate a 512×512 oxygenation map. This is 8 times faster than computing optical properties with two GANPOP networks and approximately 10 times faster than two GANPOP inferences followed by a Beer-Lambert fitting step. We expect OxyGAN to process 1024×1024 images at a similar framerate (25Hz) on a quad-GPU workstation.

To evaluate model performance, we benchmarked OxyGAN by comparing to a single-snapshot technique based on a physical model (SSOP). Table 3-I shows that, in estimating both uncorrected and profile-corrected oxygenation, OxyGAN achieves

higher accuracy than SSOP in all tissue categories. In addition to improved average accuracy, OxyGAN results also contain fewer subjective image artifacts (Fig. 3-4). These benefits are more pronounced for samples with complex surface topography, such as the pig gastrointestinal sample. Unlike SSOP, which relies on Fourier domain filtering, OxyGAN utilizes both local and high-level features. As a content-aware, data-driven approach, OxyGAN has the potential to learn the underlying distribution of the data and accurately infer oxygenation in regions with low signal or non-uniform surface structures.

In Table 3-I, we observe that GANPOP achieves similar accuracy to SSOP for profile-corrected ground truth. This is expected for several reasons. First, the training set used in this study is smaller than in the original GANPOP paper [3], excluding *in vivo* hands and tissue-mimicking phantoms. Second, for physical model-based techniques, such as SSOP, the optical property errors due to surface topography variation are correlated across wavelengths, and can later be reduced by chromophore fitting. For instance, for surface normal vectors pointing further away from the detector, the predicted absorption coefficients will be overestimated for both 659nm and 851nm. However, the fitting of oxygen saturation, which relies on the ratios of absorption coefficients, may mask the intermediate optical property errors. Because the GANPOP networks are trained independently for 659nm and 851nm, the loss function does not learn these correlations, resulting in smaller improvements in accuracy over SSOP for StO_2 measurements than for optical property measurements. This observation also provides some intuition for why the OxyGAN network might improve accuracy

over GANPOP. Because OxyGAN is trained on multi-wavelength input and the loss function is computed from the StO_2 estimate, it is capable of modeling correlations between absorption at different wavelengths and learning to reduce the effects of varying surface topography. Furthermore, a higher error rate is observed for the pig colon sample for both GANPOP and OxyGAN, likely due to this tissue type not being included in the training set and the complex topography of the colon specimen compared to other training samples.

The architecture of OxyGAN is based on the GANPOP framework [3]. The generator combines the features of both the U-Net and the ResNet, in that it incorporates both short and long skip connections and is fully residual. As discussed in Ref. [3], this fusion generator has advantages over most other existing architectures because it allows information flow both within and between levels, which is important for the task of optical property prediction. In this study, we empirically trained a model with a standard U-Net generator. The model performed well on sample types included in the training set; however, it collapsed and was unable to produce accurate results when tested on unseen sample types, such as human hands. Compared to GANPOP, OxyGAN employs data augmentation in the form of horizontal and vertical flipping, which is important for preventing overfitting of models trained on small datasets. OxyGAN also utilizes label smoothing in training the discriminator, which further improves model performance and overall training stability. Lastly, we found that adding a channel of checkerboard reference phantom measurements to the 2-wavelength structured light inputs improves accuracy for measurements taken on different days,

allowing OxyGAN to take system drift into account similarly to conventional SFDI.

In the future, more work could be done to optimize the algorithm of OxyGAN to further improve the data processing speed. The model could be trained and tested on larger datasets that span a wider range of tissue types or scenarios that might be encountered clinically. To develop a robust and generalizable model, future work should train on data with a range of spatial frequencies acquired on several different instruments. Domain adaptation techniques could also be implemented on the trained models to improve robustness to different imaging geometries. In addition, similar to other single-snapshot techniques, one limit of OxyGAN is its expected sensitivity to ambient light. Moreover, oxygenation mapping using SFDI structured illumination is currently limited because it has a shallow depth of field and requires precisely controlled imaging geometry, making its clinical adoption particularly challenging. One alternative is to use random laser speckle patterns as structured illumination, which could be less costly than SFDI projection systems, more easily incorporated into endoscopic applications, and may avoid fringe artifacts due to sinusoidal illumination [53]. Monocular depth estimation could also be incorporated for profile-correction without requiring a projector and profilometry [117, 76]. Furthermore, a more sophisticated LUT could be developed to directly estimate StO_2 from SSOP data, which models the correlations between reflectance measurements at different wavelengths and the underlying tissue oxygenation. This pixel-wise estimation may provide a more accurate baseline that will help quantify the benefit of the content-aware aspect of OxyGAN. Lastly, data-driven methods may be useful for taking higher-order

optical property effects into account, such as the scattering phase function.

3.5 Conclusion

In this study, we have presented an end-to-end approach for wide-field tissue oxygenation mapping from single structured illumination images using conditional generative adversarial networks (OxyGAN). Compared to both uncorrected and profile-corrected SFDI ground truth, OxyGAN achieves a higher accuracy than model-based SSOP. It also demonstrates improved accuracy and faster computation than two GANPOP networks that first estimate optical absorption. This technique has the potential to be incorporated into many clinical applications for real-time, accurate tissue oxygenation measurements over a large field of view.

Chapter 4

Speckle Illumination Spatial Frequency Domain Imaging for Projector-free Optical Property Mapping

This chapter contains work originally published in [53]:

4.1 Introduction

As discussed in previous chapters, optical properties provide useful information on tissue composition, oxygenation, and metabolism [55, 16, 17]. Absorption and scattering maps are being used for an increasing variety of clinical applications, including image-guided surgery, wound monitoring, and assessment of surgical margins [97, 56, 15]. Moreover, unlike color values, optical properties are absolute measurements that can be directly compared across different imaging platforms, study sites, and time-scales, facilitating their statistical interpretation via machine learning.

Over the last decade, SFDI has emerged as a powerful tool for optical property measurements. The principles of SFDI have been described in Chapter 1. In brief, conventional SFDI acquires images of tissue under sinusoidal illumination at different spatial frequencies and phase offsets, sampling the modulation transfer function (MTF) of the tissue [21]. Because the bulk tissue scattering and absorption properties preferentially attenuate the high- and low-spatial frequencies of the MTF, respectively, sampling the MTF at a range of spatial frequencies allows decoupling the two effects. Images are subsequently demodulated and calibrated, and model inversion is performed using a lookup table (LUT). SFDI can rapidly generate wide-field optical property maps using a camera and projector in a non-contact configuration. These advantages have led to the exploration of SFDI for a number of clinical and research applications [97, 56, 102, 103, 55, 15].

Despite its advantages, there are several practical challenges to translating SFDI to clinical settings. To generate an optical property map, conventional SFDI requires

a minimum of six images at each wavelength (three phase offsets at two different spatial frequencies) and performs a pixel-wise LUT search. Progress has been made towards improving the acquisition and processing time. For example, the acquisition requirements can be relaxed by estimating the MTF from a single spatial frequency using signal processing [38] and content-aware machine learning [3]. Processing speed has also been improved with GPU-based implementations and machine learning [46, 48, 49]. A second major challenge is that SFDI requires the projection of structured illumination with precisely controlled imaging geometry and known spatial frequencies. This requirement has made the translation of SFDI to endoscopy and other space-constrained applications particularly difficult. Previous work has implemented SFDI in custom benchtop endoscopy systems with an added projection channel [58, 118], but this approach requires either rigid relay optics or low-pixel-count fiber imaging bundles, each of which take up significant cross-sectional area of the endoscope.

An alternative to projecting images of patterns to sample the MTF is to utilize the speckle patterns formed by the interference of coherent illumination. Analyzing tissue response to speckle patterns is often used to estimate blood flow in laser speckle contrast imaging [119, 120]. This approach has the important advantage of being amenable to low-cost and compact implementation, for example, by laser illumination coupled through an optical fiber. Moreover, compared to image projection, coherent illumination can produce patterns with high spatial frequencies over a large depth of field. Recently, Jain et al. [121] analyzed the response of turbid liquid phantoms to laser illumination and found that speckle patterns are blurred in agreement with a model

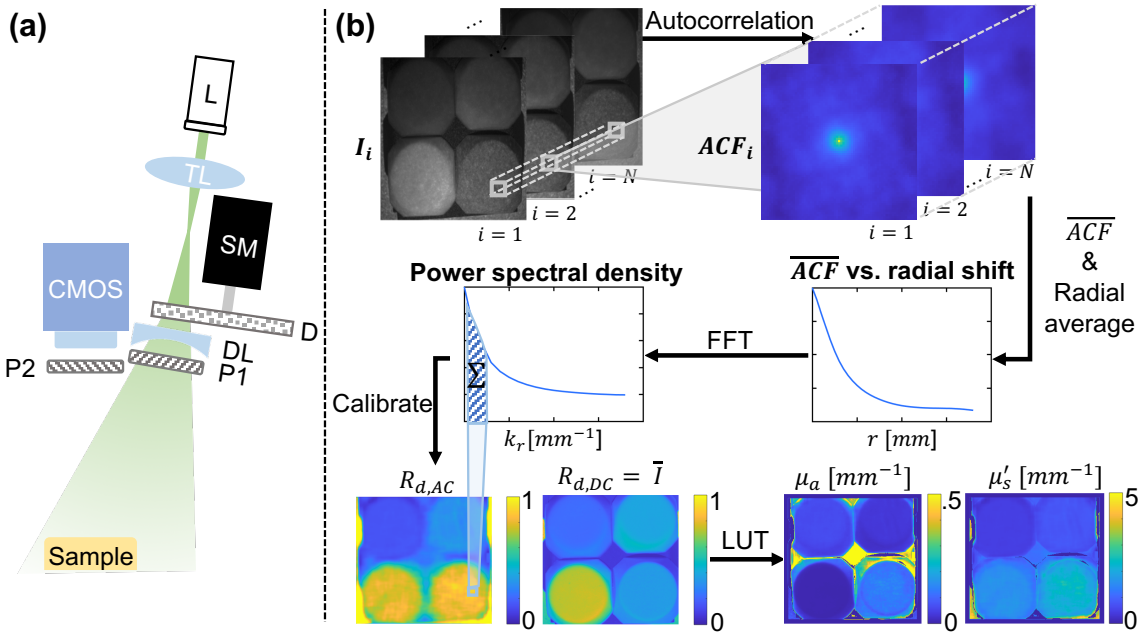

Figure 4-1. Overview of si-SFDI. (a) Experimental setup. A laser diode (L) is focused by a tunable lens (TL) on a rotating diffuser (D) mounted on a stepper motor (SM). A diverging lens (DL) spreads the light to match imaging vergence. Two linear cross polarizers (P1 and P2) reduce specular reflections. (b) Processing flow. The mean autocorrelation function (ACF) is calculated on a sliding window for N speckle images. Results are radially averaged and the Fast Fourier Transform (FFT) is taken to produce a local power spectral density (PSD). After calibration, AC and DC reflectances ($R_{d,AC}$ and $R_{d,DC}$) are used to estimate optical properties from a lookup table (LUT).

based on the optical properties of the sample. However, this technique has several limitations. First, this work developed a forward model that correlates with sample measurements but did not directly measure optical properties or evaluate accuracy. Second, this model analyzes full images, producing one response measurement for an entire image, and thus is not capable of producing an optical property map and only works on homogeneous samples.

In this study, we present speckle illumination spatial frequency domain imaging (si-SFDI), which maps the optical properties of turbid media from unknown laser speckle patterns. For the same imaging geometry and optical properties, the integral of the

power spectral density (PSD) should be constant for any speckle pattern randomization. Therefore, for the same imaging geometry but different optical properties, relative changes in the MTF can be sampled at consistent spatial frequencies via a phantom calibration, but without knowledge of the exact speckle pattern. This phenomenon can be exploited to reconstruct tissue optical property maps using as few as one speckle pattern. There are two main contributions of this study. First, we develop an inverse model to accurately measure tissue optical properties from random speckle images. Second, we apply this technique to heterogeneous, biological tissues and evaluate its performance. To our knowledge, this is the first attempt to directly map optical properties from unknown laser speckle patterns.

4.2 Methods

The experimental setup is shown in Fig. 4-1(a). A 520 nm wavelength laser diode is used for illumination (Opt Lasers micro RGB laser module) and a 1392 x 1040 pixel CMOS sensor with a camera lens ($f/\# = 2.1$) is used for imaging, resulting in 138 μm pixels in the object space when the sample is 30 cm away. A tunable lens (Optotune EL-10-30-TC) focuses the laser beam onto a 220-grit diffuser (Thorlabs DG20-220). By varying focal length, the tunable lens alters the laser spot size on the diffuser surface and subsequently, the size of the speckles in the object space. Speckle size was adjusted to be an average of 3 pixels per speckle grain width to satisfy the Nyquist criterion. The diffuser is mounted on a NEMA 11 stepper motor, which rotates at random angles to vary the speckle pattern. The light is spread by a diverging

lens (Thorlabs LD1613-A) to match the 60°camera field of view. Additionally, the illumination and detection paths are cross-polarized to reduce specular reflections. The laser illumination path is mounted at a 30-degree angle on the side of a commercial SFDI imaging head (Modulim Reflect RS^{TM}), which has a projector at a 12°angle and was used for measuring ground truth optical properties at 526 nm using 3-phase 0 and $0.2mm^{-1}$ sinusoidal illumination.

Figure 4-1(b) summarizes the si-SFDI algorithm. Laser speckle illumination, which is randomized by object and diffuser movements, is a wide-sense stationary process [121–123] with a constant autocorrelation function (ACF) and power spectral density (PSD). We characterize the tissue response in the spatial frequency domain by analyzing the PSD of each speckle image. This allows tissue optical properties to be calculated without knowledge of the exact illumination pattern.

For each speckle pattern I_i, we first apply a median filter with a kernel size of 3 pixels to reduce shot noise. The DC parameter is then estimated as:

$$M_{DC} = \frac{\sum_{i=1}^{N} I_i}{N}, \tag{4.1}$$

which is smoothed using another median filter to generate an M_{DC} map. In our case, we chose a kernel size of 7 pixels, which was the average distance between speckle grains. The estimated M_{DC} is subtracted from each I_i to isolate tissue response to spatially varying speckle patterns. We subsequently iterate through each image with a step size of 11 pixels to extract 73×73-pixel sliding windows ($w_i(x, y)$). This window size corresponds to a frequency resolution of 0.2 mm^{-1}. The lowest non-zero spatial

frequency resolvable by the sliding window is inversely related to the window size. To achieve a finer spatial frequency resolution, a larger window would be needed, but this would come at the expense of a lower spatial resolution of the reconstructed optical property maps. The step size of the sliding window depends on the desired resolution and computational resources. For samples with high levels of heterogeneity, a smaller step size may be preferred for higher resolution, at the cost of a longer computational time. The autocorrelation function (ACF) of each window is calculated using the Wiener-Khinchin theorem:

$$a_{w_i}(x, y) = \mathcal{F}^{-1}[W_i^*(k_x, k_y) \cdot W_i(k_x, k_y)], \tag{4.2}$$

where $W_i(k_x, k_y)$ and $W_i^*(k_x, k_y)$ stand for the Fourier transform of the extracted window and its complex conjugate, respectively. For each window, we compute the mean of the ACFs across N speckle patterns:

$$a_w(x, y) = \frac{\sum_{i=1}^{N} a_{w_i}(x, y)}{N}, \tag{4.3}$$

and radially average $a_w(x, y)$ to produce an ACF curve ($a_w(r)$). The power spectral density (PSD) is then computed as its Fourier transform:

$$S_W(k_r) = \mathcal{F}[a_w(r)]. \tag{4.4}$$

We define the high spatial frequency response parameter as:

$$M_{AC} = \int_{0.1}^{0.5} \sqrt{S_W(k_r)} dk_r, \tag{4.5}$$

which is the sum of the frequency response corresponding to the first two non-zero points of the PSD ($0.1mm^{-1} \leq k_r \leq 0.5mm^{-1}$).

After repeating the same steps for a reference phantom with known optical properties, the AC and DC diffuse reflectance (R_d) of the new sample can be calculated as Eq. 1.21. In this calibration, $R_{d,ref,pred}$ denotes the diffuse reflectance of the reference phantom predicted by Monte-Carlo simulations [21]. Finally, the optical properties can be found using a pixel-wise lookup table search that correlates (μ_a, μ_s') with ($R_{d,DC}$, $R_{d,AC}$). The LUT is generated by performing Monte-Carlo simulations of $R_{d,AC}$ at $0.1mm^{-1}$ increments and summing them together to match the range of frequencies sampled by the first two non-zero points of the measured PSD centered at $0.2mm^{-1}$ to $0.4mm^{-1}$, which represent the response from $0.1mm^{-1}$ to $0.5mm^{-1}$. The frequency resolution and the lowest non-DC spatial frequency are affected by the window size. For the window size used in this proof-of-concept study, the PSD frequency resolution is $0.2mm^{-1}$. We assume that each point in the PSD curve represents the integrated response over a band of frequencies. For example, the first non-zero point of the PSD is centered at $0.2mm^{-1}$ and represents the integrated response of the frequencies between $0.1mm^{-1}$ and $0.3mm^{-1}$. We took this into account when constructing the LUT with Monte-Carlo simulations by conducting these simulations at smaller spatial frequency increments than our PSD resolution. For example, to find the response

at $0.2mm^{-1}$, we simulated the reflectance at $0.1mm^{-1}$, $0.2mm^{-1}$, and $0.3mm^{-1}$ and used the average of these responses. The same process is performed for the calibration phantom. The data flow is also summarized in Algorithm 1. For comparison, the conventional SFDI processing flow is shown in Algorithm 2.

Algorithm 1 si-SFDI algorithm

procedure SI-SFDI(x, y, r, ss, N, I_i)
  average N patterns (I_i) and smooth to obtain M_{DC}
  subtract M_{DC} from each I_i
  while $(x-d, y-d) > (0,0)$ and $(x+d, y+d) \leq size(I_i)$ **do** ▷ d is window radius
    for $n = 1 : N$ **do**
      extract window ($w_i(x,y)$) with width $= 2d$
      calculate autocorrelation $a_{w_i}(x,y)$
    end for
    $a_w(x,y) \leftarrow \frac{1}{N} \cdot \sum_{i=1}^{N} a_{w_i}(x,y)$
    radially average $a_w(x,y)$ to obtain $a_w(r)$
    $S_W(k_r) \leftarrow \mathcal{F}[a_w(r)]$ ▷ $S_W(k_r)$ is power spectral density
    $M_{AC}(x - \frac{ss}{2} : x + \frac{ss}{2}, y - \frac{ss}{2} : y + \frac{ss}{2}) \leftarrow \int_{0.1}^{0.5} \sqrt{S_W(k_r)} dk_r$ ▷ ss is step size
    $x \leftarrow x + ss$
    $y \leftarrow y + ss$
  end while
  repeat above steps for reference phantom
  calibrate to obtain $R_{d,DC}$ and $R_{d,AC}$
  fit (μ_a, μ_s') using an LUT
  return μ_a and μ_s'
end procedure

4.3 Results

4.3.1 Homogeneous Phantoms

We first validated si-SFDI on homogeneous samples by imaging 16 tissue-mimicking phantoms with unique combinations of optical properties. The phantoms were fab-

Algorithm 2 Conventional SFDI algorithm

procedure SI-SFDI($I_0(x_i, f_x)$, $I_{\frac{2}{3}\pi}(x_i, f_x)$, $I_{\frac{4}{3}\pi}(x_i, f_x)$)

$M(x_i, f_x) \leftarrow$
$\frac{\sqrt{2}}{3}\times$
$\sqrt{[I_0(x_i, f_x) - I_{\frac{2}{3}\pi}(x_i, f_x)]^2 + [I_{\frac{2}{3}\pi}(x_i, f_x) - I_{\frac{4}{3}\pi}(x_i, f_x)]^2 + [I_{\frac{4}{3}\pi}(x_i, f_x) - I_0(x_i, f_x)]^2}$

▷ x_i is pixel location; f_x is spatial frequency; I_0, $I_{\frac{2}{3}\pi}$, and $I_{\frac{4}{3}\pi}$ are images at three phases

$M_{DC}(x_i) \leftarrow M(x_i, 0)$

$M_{AC}(x_i) \leftarrow M(x_i, 0.2mm^{-1})$

repeat above steps for reference phantom

calibrate to obtain $R_{d,DC}$ and $R_{d,AC}$

fit (μ_a, μ_s') using an LUT

return μ_a and μ_s'

end procedure

ricated by polydimethylsiloxane (PDMS) with titanium dioxide and India ink as scattering and absorption agents, respectively. For each phantom, we compared optical properties estimated by si-SFDI to those computed by conventional SFDI in a central 150×150 pixel region of interest (ROI). The mismatch in illumination wavelengths (526nm for SFDI and 520nm for si-SFDI) was first corrected using the predicted optical properties of India ink and titanium dioxide. The expected absorption was calculated as:

$$\mu_{a,si-SFDI} = \mu_{a,SFDI,meas}\frac{\epsilon(520nm)}{\epsilon(526nm)} = 1.013\mu_{a,SFDI,meas}, \tag{4.6}$$

where ϵ is the extinction coefficient of India ink at the corresponding wavelengths [124]. A Rayleigh scattering relationship ($\mu_s' \propto \lambda^{-4}$) was assumed for μ_s' as the diameter of titanium dioxide nanoparticles was much smaller than the wavelengths

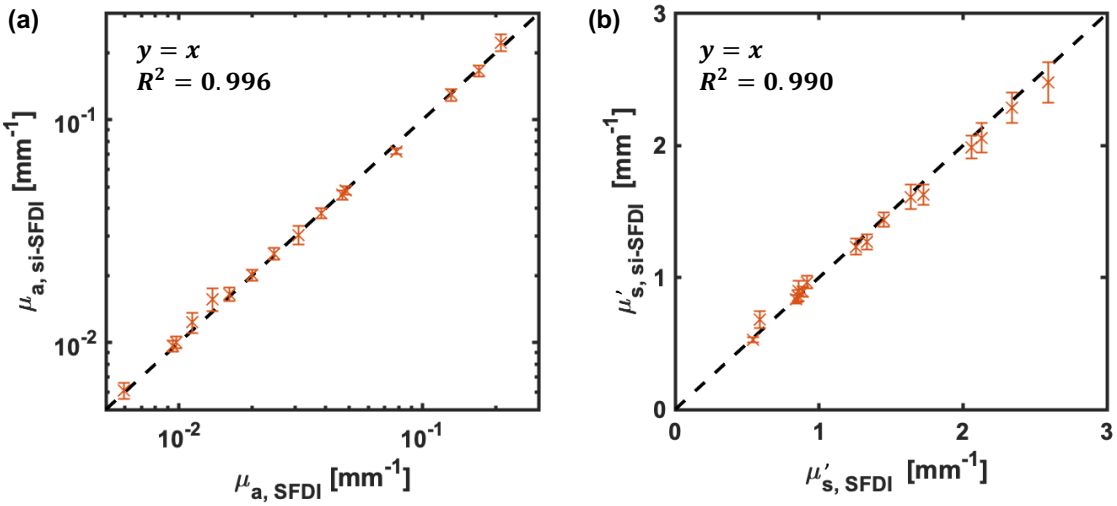

Figure 4-2. Optical property measurements from si-SFDI on 16 homogenous tissue phantoms. (a) si-SFDI measurements of absorption and (b) si-SFDI measurements of reduced scattering versus conventional SFDI ground truth.

used in this study. Thus, the expected scattering coefficient was defined as

$$\mu'_{s,si-SFDI} = \mu'_{s,SFDI,meas}\frac{(520nm)^{-4}}{(526nm)^{-4}} = 1.047\mu'_{s,SFDI,meas}. \tag{4.7}$$

With four input images ($N = 4$), si-SFDI results are plotted against SFDI ground truth in Fig. 4-2. We found that the si-SFDI predicted optical properties fit a $y = x$ ground truth curve with an R^2 value of 0.996 for absorption and 0.991 for reduced scattering. The variance of the si-SFDI measurements increases for larger μ'_s, likely due to decreased blurring of the speckle patterns, making the measurements more sensitive to local speckle grains. Analyzing the same ROI across all 16 phantoms, we observed an average pixel error of 6.9% for μ_a and 5.7% for μ'_s when using four speckle patterns ($N = 4$). Using a single speckle pattern, the average pixel errors increased to 12.9% for μ_a and 11.0% for μ'_s.

The results of si-SFDI applied on tissue-mimicking phantoms using one and four

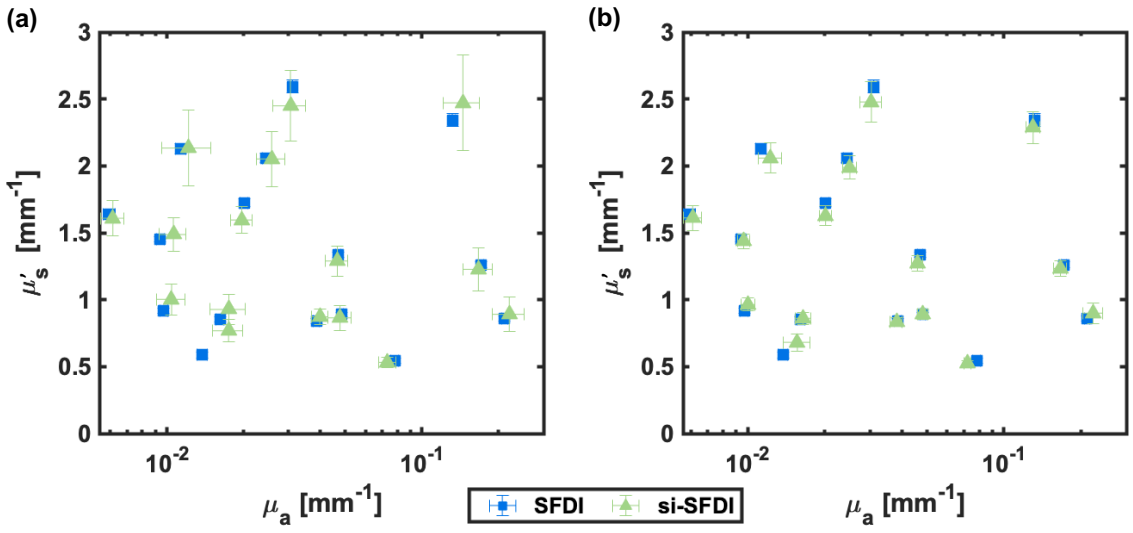

Figure 4-3. Scatter plots of phantom optical properties measured by si-SFDI. (a) Results using one speckle pattern. (b) Results using four speckle patterns. SFDI ground truth is shown in both plots. Markers indicate mean values and error bars indicate variance over a central 150×150 pixel region of interest.

speckle images are shown in 2-D scatter plots in Figure 4-3. Compared to one speckle pattern (Figure 4-3(a)), si-SFDI with four patterns (Figure 4-3(b)) shows improved accuracy with lower variances. This is expected since individual speckle patterns are often noisy and averaging multiple patterns improves the confidence in ACF estimations.

Representative PSD plots are also shown in Fig. 4-4 for five homogeneous phantoms with different optical property pairs. Expected results simulated by Monte-Carlo models are also shown in dashed lines. Our measurements demonstrate strong agreement with simulations, especially for spatial frequencies up to $0.4mm^{-1}$. As expected, the shape of the curves is determined by both μ_a and μ_s'. We observe that the measured PSDs tend to underestimate expected curves. This could be caused by the image filtering described in the si-SFDI processing pipeline, which improves the accuracy of the results but leads to errors at high spatial frequencies.

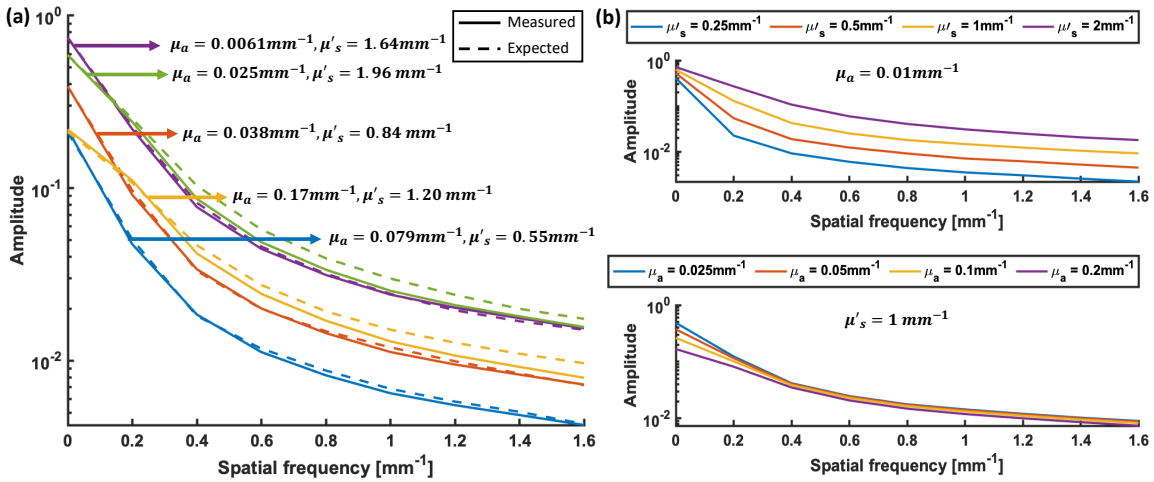

Figure 4-4. (a) Representative power spectral density plots for 5 homogeneous phantoms with varying absorption and reduced scattering coefficients. Measured results are plotted in solid lines and simulated data in dashed lines. (b) Top: Simulated PSD plots with fixed μ_a and varying μ'_s; bottom: simulated PSD plots with fixed μ'_s and varying μ_a.

4.3.2 Heterogeneous Tissue Samples

To assess the accuracy of si-SFDI in heterogeneous samples with irregular surface topography, we tested the technique on *in vivo* human hand and *ex vivo* swine gastroesophageal junction (Fig. 4-5(a), (b)). Side-by-side comparisons of 4-image si-SFDI with ground truth SFDI show that for flat regions, there is excellent agreement over the whole field of view. si-SFDI recorded a sharp change in optical properties between the stomach and esophagus regions of the swine sample, in agreement with SFDI, and representative of the expected change in tissue type on either side of the junction.

To explore the effect of the number of speckle patterns on the accuracy of optical property calculations, we analyzed a relatively-flat region on the back of the hand and at the gastroesophageal junction (250×250 pixels, highlighted by the white boxes in Fig. 4-5(a) and (b)) and plotted the error with varying numbers of images (Fig.

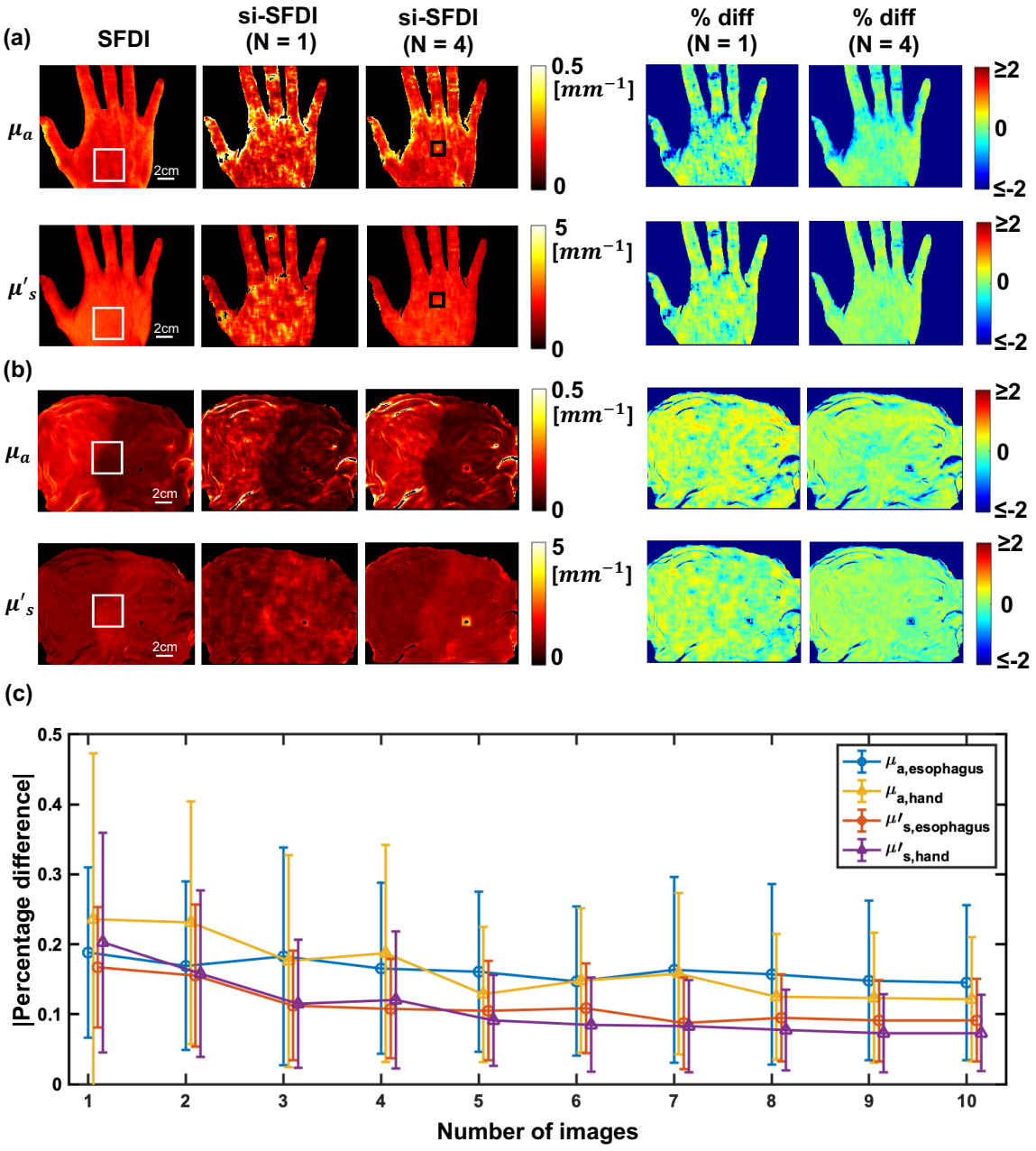

Figure 4-5. Comparison of 1- and 4-image si-SFDI optical property estimates with SFDI ground truth in (a) *in vivo* human hand, and (b) *ex vivo* pig gastroesophageal junction. N indicates the number of input patterns. (c) Absolute percentage difference as a function of number of speckle patterns used in the si-SFDI calculation for the ROI indicated by the white boxes in (a) and (b). Error bars are standard deviations. Black boxes in (a) highlight a scar that did not produce contrast in SFDI.

4-5(c)). This error was calculated as the absolute percentage difference. The errors for both absorption and scattering measurements decrease significantly when additional speckle patterns are included, but beyond $N = 4$ the improvements become small–the decrease in errors from $N = 4$ to $N = 10$ images is only approximately 2%. Analyzing additional speckle patterns has the effect of both increasing signal to noise and improving the spatial resolution of optical property mapping, since additional patterns improve the sampling of the local ACF in our window-based approach. Single-image si-SFDI results ($N = 1$) are also shown in Fig. 4-5(a) and (b). As expected, the results show fair agreement with SFDI, however, with more noise and image artifacts than $N = 4$. To further investigate the accuracy and image quality of si-SFDI applied to heterogeneous samples, we calculate the structured similarity (SSIM) index for the pig esophagus and human hand sample (Table 4-I). Overall, si-SFDI demonstrates high SSIM scores compared to profile-corrected SFDI as reference. On average, si-SFDI with 4 images ($N = 4$) achieves 6.3% higher SSIM than $N = 1$, and this improvement becomes 8.2% with $N = 10$.

For both hand and esophagus samples (Fig. 4-5), we observe a larger discrepancy between si-SFDI and ground truth absorption than scattering measurements. This may be partially due to the mismatch in wavelengths used (520nm and 526nm for si-SFDI and SFDI, respectively). For example, the extinction coefficients of hemoglobin are different at these two wavelengths, with a ratio of approximately 1 to 1.2 [125]. This mismatch could also contribute to the differences in scattering, but likely to a lesser extent [126]. Moreover, in areas of the sample with irregular surface topography, we

Table 4-I. SSIM of pig esophagus and human hand optical properties obtained by si-SFDI. SFDI is used as reference.

Number of speckle patterns (N)	Pig esophagus		Human hand	
	μ_a	μ_s'	μ_a	μ_s'
$N = 1$	0.790	0.870	0.777	0.817
$N = 4$	0.839	0.901	0.846	0.871
$N = 10$	0.855	0.912	0.874	0.893

found larger errors in absorption than scattering estimates. This is expected because absorption depends mostly on M_{DC}, while scattering depends mostly on M_{AC}, which is less affected by surface angle variations. Additionally, the difference in illumination angle for the laser diode compared to the SFDI projector is expected to contribute to reconstruction errors.

To explore the effect of surface topography on the accuracy of si-SFDI, we compare the absolute percentage difference (Fig. 4-6(a) for μ_a and (b) for μ_s') with surface normal angles (Fig. 4-6(c)) of the pig esophagus sample. Qualitatively, regions with larger angles are associated with higher error rates. This is expected since the ground truth is profile-corrected SFDI while si-SFDI does not account for surface topography. We additionally plotted the normalized mean absolute error (NMAE) with respect to surface normal angles (Fig. 4-6(e)). The mean error of si-SFDI is below 20% for relatively flat regions (<10 degrees) and increases with larger tissue angles. Moreover, absorption coefficients appear to be more affected by surface normal angles than reduced scattering, which corresponds to findings in previously published studies [30, 34] and may additionally be due to the window-based approach for calculating M_{AC}. Moreover, theoretically, because of the broader bandwidth of spatial frequencies

contained in the illumination pattern, si-SFDI is expected to sample optical properties from a larger axial range than SFDI. Although this was not assessed here, it may warrant further investigation in future studies.

We observed a small region with large errors on the hand sample (highlighted by the black boxes in Fig. 4-5(a)). From inspection of the hand, we found a scar in this region that does not generate contrast in conventional SFDI but was captured with si-SFDI. This difference in contrast may be due to the si-SFDI incorporating higher spatial frequencies than SFDI, which would lead to a shallower sampling of tissue optical properties [21] and also contributions from sub-diffuse scattering parameters. A recent study similarly uncovered a scar when using high spatial frequency structured illumination [127].

4.3.3 Effect of Blood Flow on Si-SFDI Optical Property Estimation

To evaluate the effect of blood flow on si-SFDI measurements, we performed an *in vivo* occlusion study where a pressure cuff was applied to an arm for 3 minutes at approximately 200mmHg. Both si-SFDI and SFDI data were acquired before and near the end of occlusion (Fig. 4-7). This study protocol was approved by Johns Hopkins Institutional Review Board.

To measure the level of blood flow, local speckle contrast (K) is computed within 5×5-pixel sliding windows as the ratio between the standard deviation and mean intensity. As expected, a higher K is observed during occlusion due to reduced blood

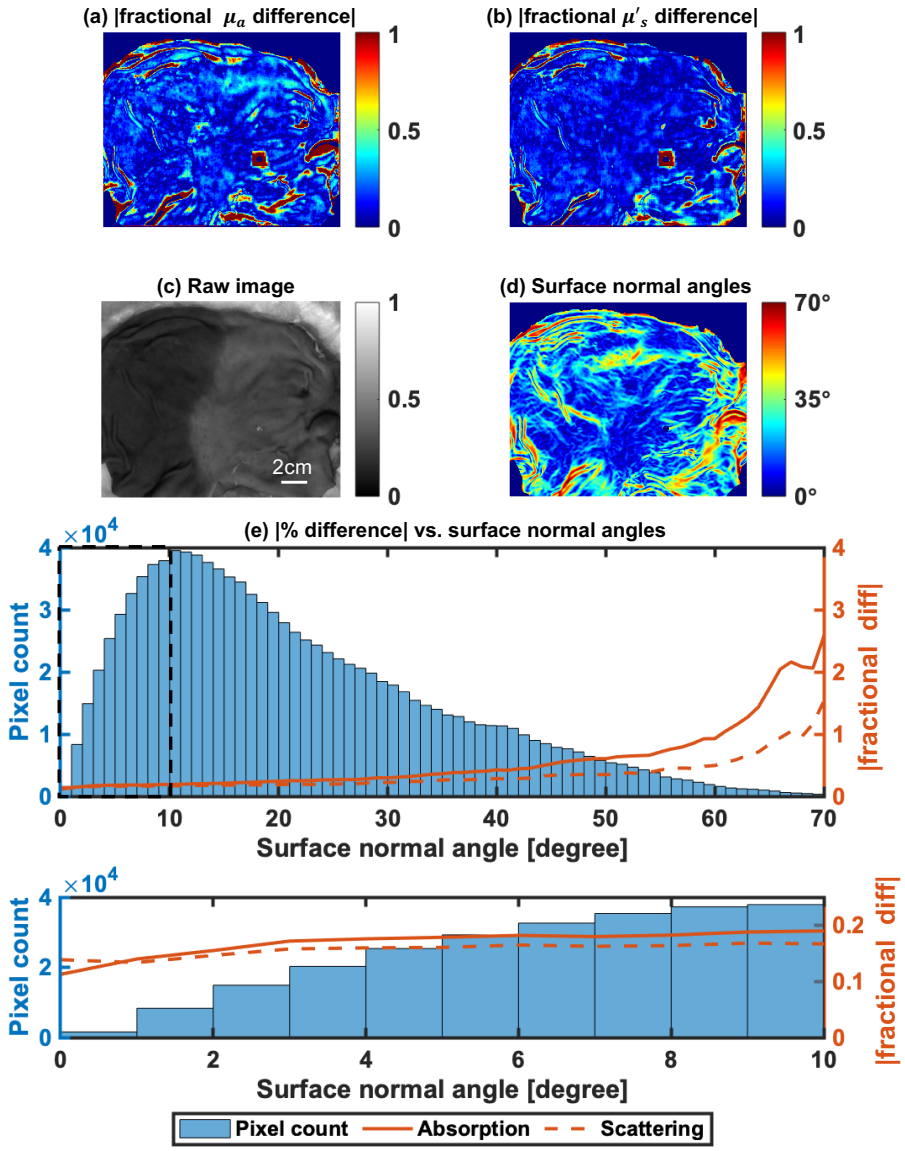

Figure 4-6. Absolute percentage difference between si-SFDI and SFDI applied to the full pig esophagus specimen. (a) Absorption error map. (b) Reduced scattering map. (c) Raw image of the sample. (d) Surface normal angles. Background is masked out in (a), (b), and (d). (e) Normalized mean absolute error (NMAE) of si-SFDI as a function of surface normal angles. si-SFDI optical properties are calculated from four input speckle patterns. Top plot in (e) shows the full range, and the region highlighted by the dashed box (between 0 and 10 degrees) is magnified in the bottom plot.

flow (Fig. 4-7(g)-(i)). The reduced scattering coefficients at baseline (resting) and occlusion measured by SFDI and si-SFDI ($N = 1$) are shown in Fig. 4-7(a)-(b) and (d)-(e), respectively. The relative change in μ_s' is shown in (c) and (f). We hypothesize that, while μ_a changes with the level of oxygenation, μ_s' is not affected by occlusion, and thus only μ_s' results are analyzed here. Despite a global increase in speckle contrast during occlusion (approximately 30%), there is not an overall change in μ_s' measured by either SFDI or si-SFDI (less than 1%). Therefore, we conclude that si-SFDI measurements are relatively insensitive to blood flow. This is due to optical properties being a function of the low-spatial frequencies, whereas flowing particles blur the high-spatial frequencies of the illumination. The si-SFDI algorithm described here analyzes the tissue response to spatial frequencies less than $0.5mm^{-1}$. To illustrate this effect, we plot the PSD over a range of spatial frequencies for an ROI with high K contrast changes due to occlusion (Fig. 4-7(j)). We observe little difference in the PSD between resting and occlusion at the low spatial frequencies used for optical property calculations. The effect of flow becomes more pronounced at spatial frequencies greater than $1mm^{-1}$, with increased attenuation of the PSD at high spatial frequencies due to resting blood flow. This result highlights the potential to image optical properties and flow contrast simultaneously by analyzing different spatial frequency ranges of the computed PSD.

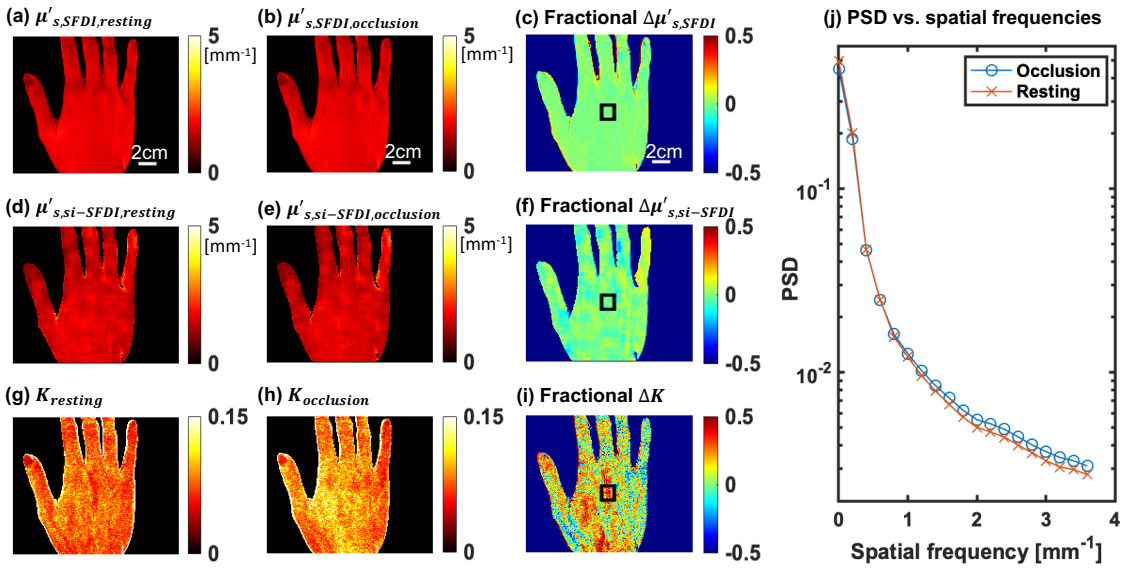

Figure 4-7. Occlusion study. (a)-(c): Resting and occlusion μ'_s maps measured by SFDI and the percentage difference; (d)-(f): Resting and occlusion μ'_s maps measured by $N = 1$ si-SFDI and the percentage difference; (g)-(i): Resting and occlusion speckle contrast maps (K) and the relative change. K is computed as the local standard deviation over mean of a 5×5 pixel sliding window; (j): PSD plots for an ROI during resting and occlusion. The ROI is highlighted by the black box in (c), (f), and (i).

4.3.4 Drift

We tested the drift of the si-SFDI system by measuring the same phantom every 10 min for 100 min. The standard deviation of the absolute variation in μ_a and μ'_s were 5.54% and 3.15%, which is similar to other SFDI drift results [128] and demonstrates stability of the system over time. An improved laser source with more stable output would further reduce system drift.

4.3.5 Step Function

To evaluate the resolution of si-SFDI, we record its response to a step function phantom and compare it with conventional SFDI (Fig. 4-8). The phantom was constructed by cutting two phantoms with different optical properties and press-fitting them together.

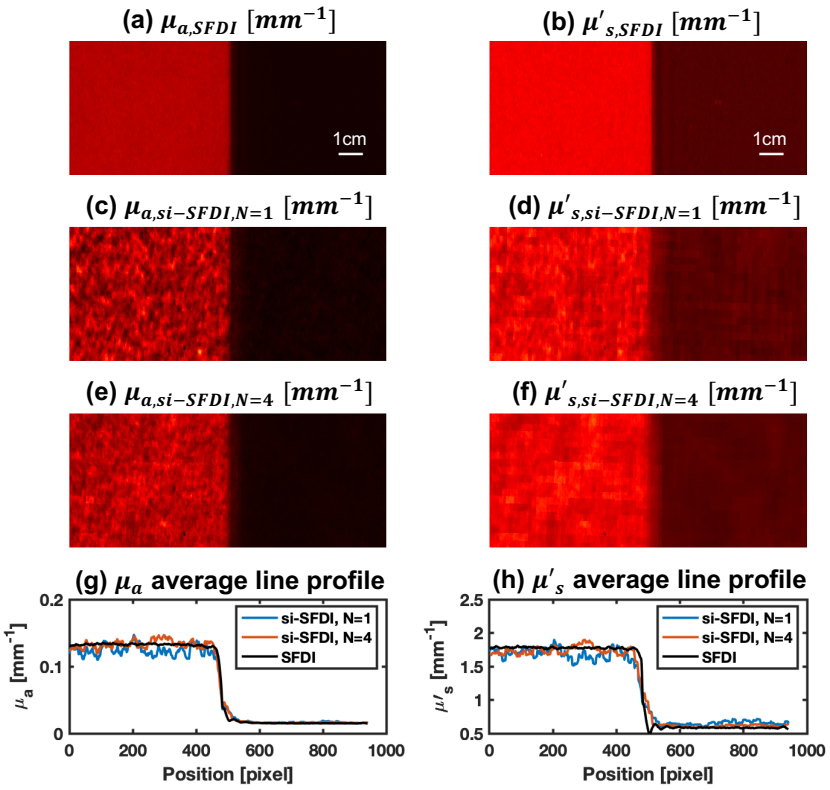

Figure 4-8. Response to a step function phantom. Figures in the left column are absorption and figures in the right column are reduced scattering coefficients. From top to bottom: conventional SFDI, si-SFDI with 1 speckle pattern, si-SFDI with 4 speckle patterns, and average line profiles.

The resolution is defined to be the distance where the contrast is reduced by 90%. For μ_a, the resolution is 4.1mm for conventional SFDI and 6.3mm for $N = 1$ si-SFDI. For μ'_s, the resolution is 2.9mm and 7.3mm for conventional SFDI and $N = 1$ si-SFDI, respectively. μ'_s resolution of si-SFDI is worse than μ_a due to windowing and a stride size larger than 1 pixel.

4.4 Discussion

In this chapter, we demonstrate a wide-field, projector-free technique for non-contact optical property mapping using laser speckle patterns as structured illumination.

There are several areas for improvement of the si-SFDI technique. For example, a source of discrepancy between SFDI and si-SFDI results is the difference between sampling depth. Theoretically, because of the broader bandwidth of spatial frequencies contained in the illumination pattern, si-SFDI is expected to sample optical properties from a larger axial range than SFDI. Although this was not assessed in this study, it may warrant further investigation in future studies.

Second, real-time acquisition and processing is imperative for incorporating si-SFDI into future clinical endoscopic applications. For the same illumination power, the exposure time to acquire a single si-SFDI image is similar to conventional SFDI. The total acquisition speed for an optical property estimate depends on the number of speckle patterns used. In the setup used for this study, a 4-image si-SFDI sequence took approximately 2 seconds to acquire. However, for constant power, this speed could be dramatically increased if the object is placed at endoscopic-distances from the camera (typically 2-5cm instead of the 25-30cm tested in our setup). Another issue arises from the fact that multiple speckle patterns are required for the most accurate optical property estimate. Although this can be achieved by using a rotating diffuser or phase randomization, it inevitably limits the imaging speed. However, instead of averaging multiple speckle patterns to approximate the DC response, one can use a laser speckle reducer to approximate planar illumination and accurately estimate M_{DC} from a single image. In this scenario, only two images would be required. Moreover, if the average illumination power of these two images is different and known, sensitivity to ambient lighting can be reduced. Alternatively, planar illumination

may be approximated from coherent illumination using a content-aware deep learning approach [129], allowing accurate single-frame si-SFDI.

In terms of processing speed, si-SFDI is currently slow due to window-based computations. It takes 2 seconds to compute a 1040×1392-pixel optical property map from a single speckle pattern ($N = 1$) using a 4-core 3.6 GHz processor running the algorithm in MATLAB (R2020a, MathWorks). However, because each window computation is independent, this computation is highly parallelizable and can be accelerated by using a graphics processing unit.

Third, for many clinical applications, the topography of the sample will need to be estimated and accounted for. This may be accomplished in simple setups with recent monocular depth estimation techniques [117, 76]. Additional research is needed to study the utility of diffuse and sub-diffuse optical signatures in endoscopic applications. We believe that si-SFDI can be applied to endoscopic measurements of tissue oxygenation by estimating the absorption coefficients at different wavelengths. Scattering coefficients could also be used to distinguish between tissue types, such as benign and malignant tissues [10].

Chapter 5

Endoscopic Imaging of Tissue Biomarkers with Speckle Illumination and Machine Learning

This chapter contains work originally published in [52]:

5.1 Introduction

In the previous chapter, we described a signal-processing approach for wide-field optical property mapping using random laser speckles, called Speckle Illumination Spatial Frequency Domain Imaging (si-SFDI). We modeled the tissue spatial MTF by analyzing the spectral response to a band of spatial frequencies, and performed calibration and optical property inversion in a similar fashion to conventional SFDI. Compared to SFDI, si-SFDI demonstrated high accuracy on homogeneous imaging phantoms and heterogeneous tissue samples. More importantly, it eliminated the need for a projector for structured illumination, making it more implementable in clinical settings.

Despite the potentials of si-SFDI, there currently exist a few limitations that need to be addressed. For example, si-SFDI suffers from slow processing speed due to the fact that it calculates the autocorrelation function on a sliding window basis. For example, it currently takes 2 seconds to process a full-resolution image using a single speckle pattern. Another issue is the presence of noise and image artifacts because of the nature of speckles and window-based computations. Moreover, it tends to produce large errors in regions with non-flat surface topography, highlighting the importance of profile corrections in estimating the most accurate optical properties.

In this chapter, we aim to incorporate si-SFDI into clinical endoscopic systems and improve its performance with machine learning. We first explore the potential of si-SFDI applied to a commercially available Olympus colonoscope. Initial results are presented and current issues are discussed. We then design a dual-channel laparoscope

system that uses stereo vision for depth estimation and rapidly generates high-quality biomarker maps enabled by deep learning. The implications of existing challenges and potential future directions are outlined.

5.2 Towards SFDI Endoscopy with Structured Illumination from Randomized Speckle Patterns

In this part of the chapter, we present the experimental setup of speckle illumination incorporated into a commercially available endoscope. We also show the preliminary results of endoscopic si-SFDI applied to homogeneous imaging phantoms and a hand during an occlusion study.

5.2.1 Methods

The workflow of si-SFDI is described in Chapter 4. In brief, because speckle illumination is a wide-sense stationary process, for the same imaging geometry, the autocorrelation function (ACF) and power spectral density (PSD) of any speckle randomization remain constant [122, 123]. The ACF can be estimated utilizing the Wiener-Khinchin theorem (Eq. 4.2). We compute the ACF for a sliding window, where the window size depends on the desired spatial frequency resolution ($0.2mm^{-1}$ in this study). The ACF at each window location is calculated as the average across N windows, where N is the number of random speckle patterns. We assume the system impulse response is radially symmetrical and express the ACF as a radially-averaged

function, $a_v(r)$. The power spectral density $S_V(k_r)$ is then calculated as the Fourier transform of $a_v(r)$. From the PSD we define two parameters, M_{AC} and M_{DC}, where M_{AC} is estimated as the area under the curve of the PSD between 0 and $0.4mm^{-1}$. M_{DC} is the response to continuous-wave illumination, which could be a single speckle-reduced image or estimated as the mean of N speckle images. We calibrate these two parameters against a reference phantom with known optical properties under coherent illumination, and estimate the diffuse reflectance (R_d) using the reference model response predicted by Monte Carlo simulations (Eq. 1.21). The estimated AC and DC reflectance can then be inverted using a lookup table to obtain μ_a and μ_s'. Overall, si-SFDI achieves high accuracy compared to conventional SFDI as ground truth.

The experimental setup of endoscopic si-SFDI is depicted in Fig. 5-1. We use a commercially available Olympus CHHQ-190L colonoscope and fiber-coupled laser sources (532 and 650 nm). The tip of the endoscope is placed 2 cm away from the sample surface. Cross-polarizers are utilized to minimize the effect of specular reflections. A laser speckle reducer (LSR-3005-24D-VIS, Optotune) toggles on and off to generate continuous-wave (DC) and speckle-illuminated (AC) images. Therefore, for each sample, one DC and one AC image are acquired. The laser diodes are synchronized with the colonoscope frame rate so that the sources may be toggled on and off at specified duty cycles to reduce the effect of uncontrolled exposure times. Prior to applying the si-SFDI algorithm, we correct the acquired images for barrel distortion and crop the central region-of-interest for analysis.

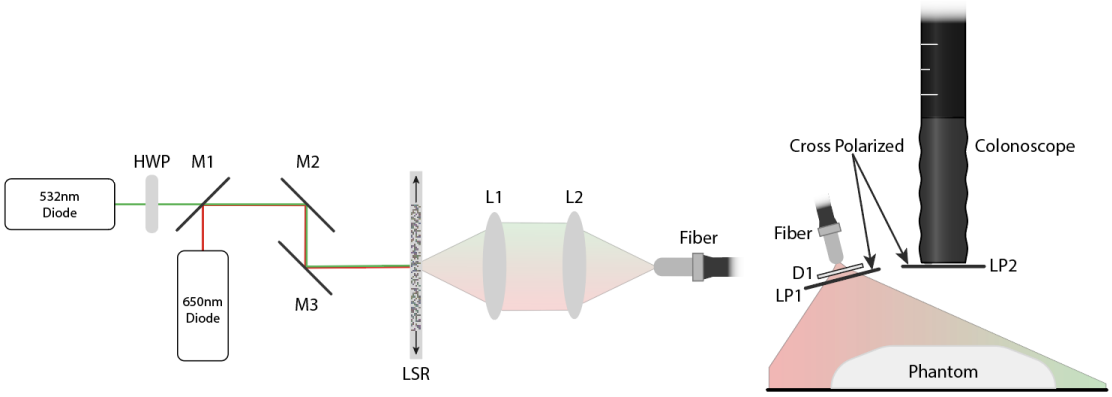

Figure 5-1. Hardware setup of endoscopic si-SFDI, which includes an Olympus 190 colonoscope and two fiber-coupled laser sources at 532 and 650 nm. Cross-polarizers are employed to mitigate specular reflections, and a laser speckle reducer is toggled on and off to produce a DC and an AC image for each sample.

5.2.2 Preliminary Results

We first apply endoscopic si-SFDI to six homogeneous tissue-mimicking phantoms. One major challenge is the lack of precise control of the endoscope exposure time, which is crucial for accurately determining diffuse reflectance. Nevertheless, we observe that the measured ratios between μ_s' and μ_a agree with the ground truth ratios obtained with conventional SFDI (Fig. 5-2(a)). Assuming a linear relationship, we find an R^2 value of 0.9812 for μ_s'/μ_a between 4 and 200.

We additionally apply endoscopic si-SFDI to the palm of a volunteer's hand during an occlusion study (Fig. 5-2). Images are acquired before and 2 minutes into the occlusion. At 660 nm illumination, we find a 53.6% increase in μ_a due to occlusion, which is expected as the extinction coefficient of de-oxygenated hemoglobin is higher than that of oxygenated hemoglobin. There is no apparent change in μ_s'. This study is approved by the Johns Hopkins Medicine Institutional Review Boards.

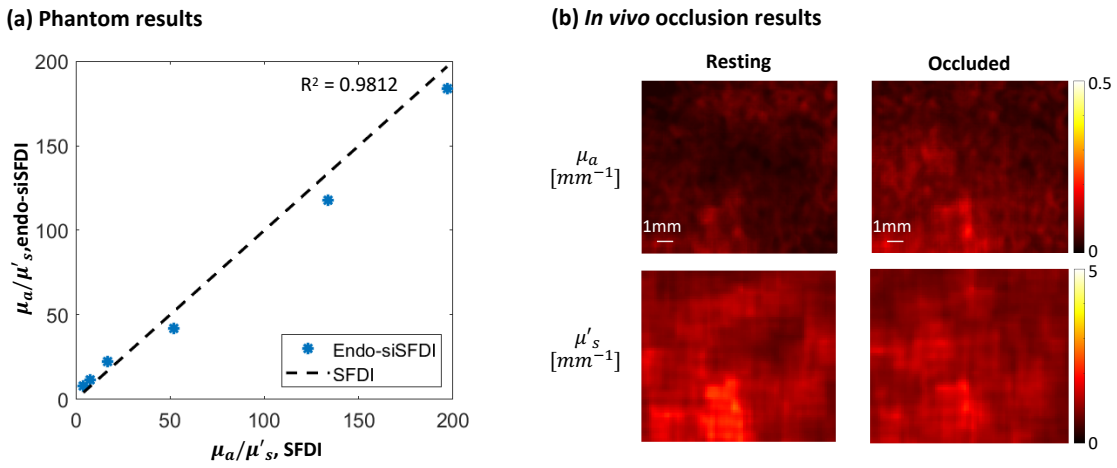

Figure 5-2. Preliminary results of endoscopic si-SFDI. (a) Comparison of ratios between μ'_s and μ_a measured by endoscopic si-SFDI and conventional SFDI. (b) μ_a and μ'_s of the palm of a hand measured by endoscopic si-SFDI before and 2 minutes into occlusion.

5.2.3 Discussion

In this section, we demonstrate the preliminary results of incorporating si-SFDI into an endoscope. We observe that endoscopic si-SFDI measures the correct ratios of absorption to scattering coefficients when applied to an *in vivo* hand. Comparing ratio results to conventional SFDI, the endoscopic si-SFDI technique agrees well in homogeneous phantoms.

There are several areas for improvement of the endoscopic si-SFDI technique. The accuracy would be improved with precise control of the endoscope exposure time and gain. Instead of externally illuminating the sample, the fiber optics could be incorporated into the colonoscope through the instrument channel or light could be coupled through the existing light guides for *in vivo* optical property mapping. Profilometry correction is also important for accurate measurements of optical properties. The surface profile could be estimated with random speckle patterns by using

either a model-based technique [130] or a deep learning approach for monocular depth estimation [117, 76]. It is possible that both conventional imaging and si-SFDI imaging can be achieved via coherent illumination, using speckle noise suppression algorithms to recover the DC image [129]. In addition, because the window-based calculation of si-SFDI is highly parallelizable, the processing speed could be improved with parallel computing and a graphics processing unit. Furthermore, machine learning models could be useful for improving the accuracy and image quality with content-aware estimation [49, 3].

5.3 Rapid Optical Property Endoscopy with Hilo Illumination and Machine Learning

In the previous section of this chapter, we explore applying si-SFDI to a clinical colonoscope. Although the initial results are promising, the lack of precise control of the camera exposure and gain through the commercial software is a major hurdle to achieving absolute measurements. As discussed in previous chapters, profilometry correction is crucial for accurate optical properties, and it is difficult to estimate sample shape with a monocular endoscope. In this section, we describe a laparoscopic system that is capable of acquiring rapid, high-fidelity optical properties with machine learning and pairs of planar/speckle illumination images. A dual-channel laparoscope is utilized to obtain depth reconstructions from stereo vision. We demonstrate preliminary results and discuss future directions that would improve the performance and clinical

utility of the system.

5.3.1 Methods

In this research, we propose a rapid endoscopic optical property mapping technique using machine learning and high-spatial frequency / low-spatial frequency ("HiLo") illumination. We modify a commercially-available dual-viewing-channel endoscope to incorporate coherent RGB laser illumination. The hardware setup is shown in Fig. 5-3. We use a da Vinci endoscope (Type S, Intuitive Surgical, Inc.) with two color cameras (1 and 2, or left and right) (GS3-U3-41C6C-C, Teledyne FLIR LLC) for stereo vision.

In si-SFDI, we use laser speckles as structured light for optical property mapping. Compared to conventional sinusoidal patterns, laser speckles have many advantages. For example, instead of a single spatial frequency, random speckles can sample a wide range of spectral content. They can also generate very high spatial frequencies at a large depth of field. More importantly, in contrast to bulky projectors, laser light can be fiber-coupled into compact systems for structured illumination. The quality or contrast of speckle grains is related to the coherence length of the laser source. Laser coherence length is defined as the optical path length difference over which a self-interfering beam maintains a certain degree of coherence, and can be expressed as:

$$L_c = \sqrt{\frac{2\ln 2}{\pi}} \frac{\lambda_0^2}{\Delta\lambda}, \tag{5.1}$$

where λ_0 and $\Delta\lambda$ are the central wavelength and the bandwidth of the source, respectively. This means that, to maintain a certain degree of speckle contrast, a diode with

a long coherence length can sustain a larger path difference than one with a short coherence length. In the context of speckle formation, the coherence of a laser beam, and hence the speckle contrast, is reduced by traveling through a longer optical fiber or being scattered/refracted by multiple rough surfaces, such as an optical diffuser.

In the setup of the HiLo endoscope (Fig. 5-3), red (636nm), green (525nm), and blue (470nm) lasers with short coherence lengths are used to produce DC (Lo) images. The three wavelengths are used for several reasons. First, in speckle-reduced mode (Lo), they can create white-light illumination for conventional endoscopic visualization. Second, we can extract tissue response to the three wavelengths in a single capture by separating R, G, and B channels of the color cameras. Third, we can use the absorbance at these three wavelengths for chromophore measurements. An additional red laser (637nm, CNI Lasers) with a long coherence length generates speckle patterns in Hi mode, and we can toggle between Hi and Lo mode by switching between the 637nm and 636nm laser. A polarizing beam splitter (PBS) is used to create a common path between the two red lasers. Laser beams are coupled into an optical fiber for illumination in conjunction with dichroic mirrors.

To acquire SFDI ground truth for supervised machine learning, we mount a projector (DLP LightCrafter, Texas Instruments) at an angle to avoid occlusion. Geometric calibration is performed to correct for the projection angle and produce a consistent spatial frequency in each pattern. Additionally, cross polarizers are employed between cameras and light sources in order to isolate diffuse signals by removing specular reflections.

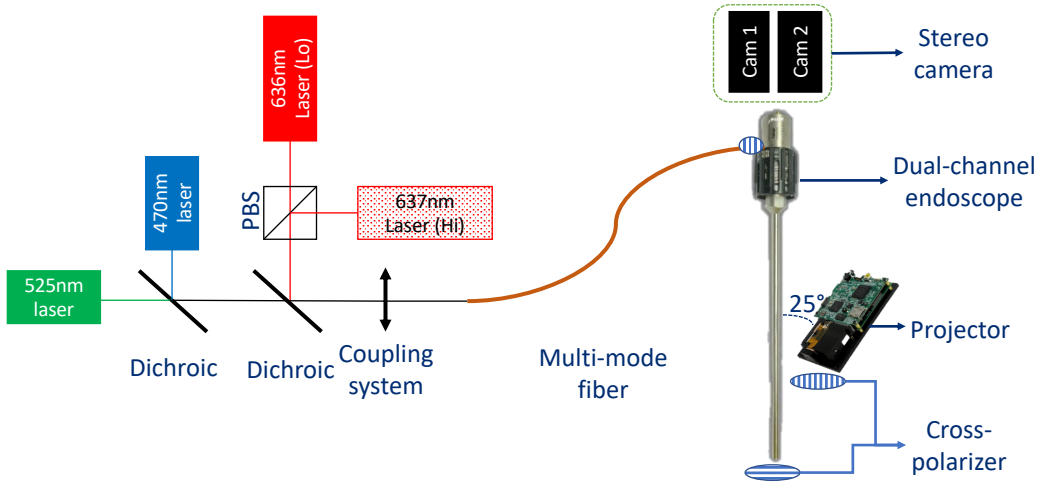

Figure 5-3. Hardware setup of the HiLo endoscope. Red (636nm), green (525nm), and blue (470nm) lasers with short coherence lengths produce conventional white-light illumination in Lo mode. Another red laser (637nm) with a long coherence length is used to generate speckle patterns in Hi mode. We toggle between Hi and Lo mode by flashing the 637nm and 636nm laser, respectively. A polarizing beam splitter (PBS) is used to create a common path between the two red lasers. The light is coupled into a multimode fiber, which is then butt-coupled into the illumination port of the laparoscope. The laparoscope is a dual-channel da Vinci endoscope that uses a stereo camera two reconstruct depth information. A projector is mounted at a 25-degree angle to acquire ground truth data using conventional SFDI. Cross polarizers are added to isolate diffuse reflection by reducing the effects of specular reflection.

To produce high-quality optical property maps in real time, we propose using a supervised machine learning model with HiLo images as input and SFDI results as ground truth. The detailed workflow is shown in Fig. 5-4. For the HiLo setup, we acquire two images per camera per sample (1 Hi and 1 Lo). Left and right images are taking simultaneously by hardware-triggering the cameras. Depth is reconstructed using either passive stereo with Lo image pairs or active structured-light stereo with Hi image pairs [131, 132]. Briefly, to obtain depth from stereo, the left and the right camera are carefully calibrated using a pattern with distinct landmarks (such as corners or circles) of known sizes. In this study, we use a 8x11 checkerboard pattern

with a square size of 2.5mm. The purpose of this calibration is to account for image distortions intrinsic to each camera and to estimate the relative pose (rotation and translation) between the two cameras. Acquired frames are first rectified using the relative pose so that corresponding features in the left and right image lie on the same horizontal line. These features are then matched with a similarity score, and the disparity is calculated, which is the pixel distance between the left and the right image for the same feature. Finally, we can compute depth from stereo following the principles of triangulation. The difference between active and passive stereo is that active stereo uses structured illumination, which improves the accuracy of feature matching, especially in homogeneous regions with little textural or color information.

Shown in Fig. 5-4, we see that machine learning is only applied to the AC response estimation, which is the most time-consuming step and the source of errors and image artifacts in si-SFDI. The algorithm is similar to the ones described in Chapters 2 and 3. DC response can be obtained by applying the calibration model directly to the Lo images. We can then apply profilometry correction to the response maps using depth from stereo and invert the profile-corrected optical properties using an ultrafast LUT [133].

For ground truth measurements, we capture three phase offset SFDI images at 0 and 0.2mm^{-1}, which is a spatial frequency pair typically used for diffuse optical property measurements [29]. It is important to note that, since only the 637nm laser produces speckles, we are only able to obtain the optical properties at the red wavelength. However, because tissue scattering coefficients monotonically decrease

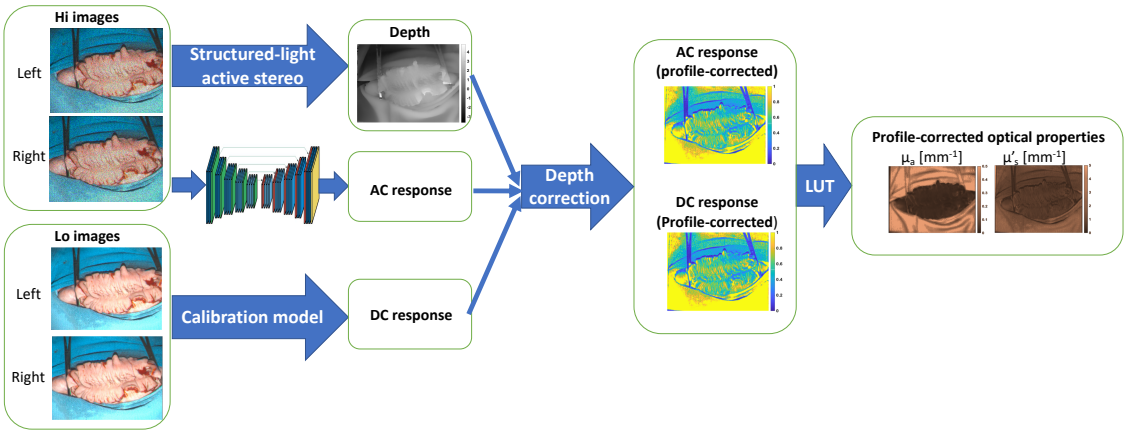

Figure 5-4. Proposed workflow for optical property endoscopy with HiLo illumination and machine learning. Left and right image pairs are used to estimate depth from either passive stereo (Lo images) or structured-light active stereo (Hi images). Machine learning is applied to bypass window-based processing of si-SFDI in order to improve speed and image quality of AC response maps. A calibration model is applied directly to Lo images to generate DC response. Profilometry correction is done on the response maps using depth from stereo, and profile-corrected optical property maps are obtained using an ultrafast lookup table (LUT).

with wavelengths in the visible regime, we can extrapolate μ_s' at the green and blue wavelengths following the Mie theory [12, 134].

To train a robust machine learning model, we ensure that the samples span a large range of optical properties, topographies, and levels of surface homogeneity. To span these ranges, we include a variety of sample types, such as tissue-mimicking imaging phantoms and *ex vivo* swine tissues, including colons and small intestines. For the preliminary results, we use a dataset that includes 19 homogeneous phantoms, 3 ex vivo swine colons, and 3 ex vivo swine intestines.

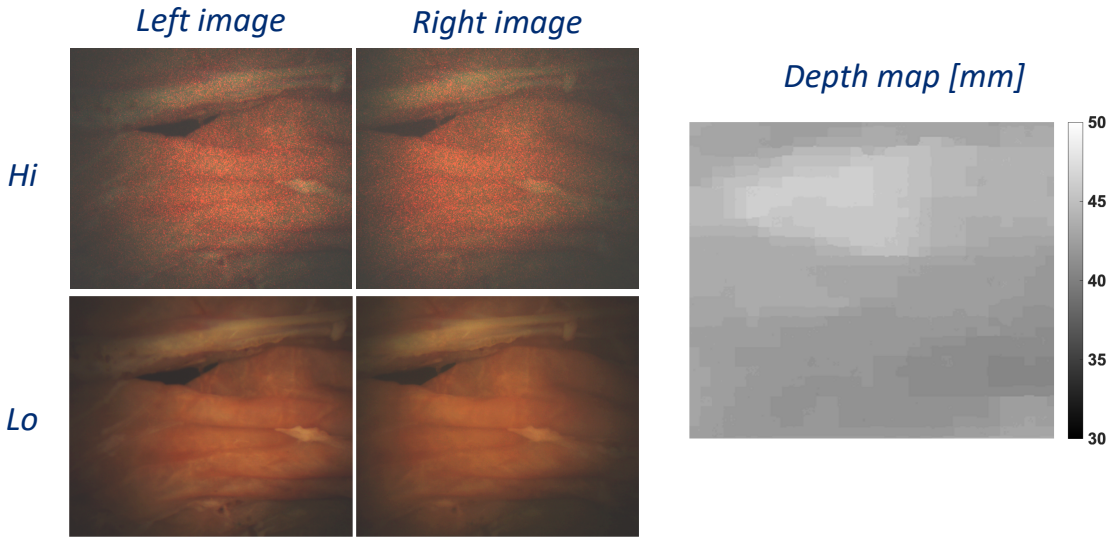

Figure 5-5. Left: example left and right image pair under Hi (top) and Lo (bottom) illumination. Right: corresponding depth map from passive stereo in millimeters.

5.3.2 Results

Example left and right images of a colon sample under Hi and Lo illumination are shown in Fig. 5-5. The corresponding depth map using the Lo image pair and passive stereo is included in the right side of the figure. We are not able to obtain accurate depth from active stereo and Hi images due to the lack of correspondences, which will be discussed in more details in the Discussion section. In spite of this limitation, Fig. 5-5 shows that state-of-the-art disparity matching techniques [135, 136] with passive stereo can still produce dense results with no data holes.

Example optical property maps from conventional SFDI (ground truth), si-SFDI, and HiLo with machine learning are shown in Fig. 5-6. It is apparent that machine learning generates results that more closely resemble the ground truth than si-SFDI in terms of accuracy and image quality.

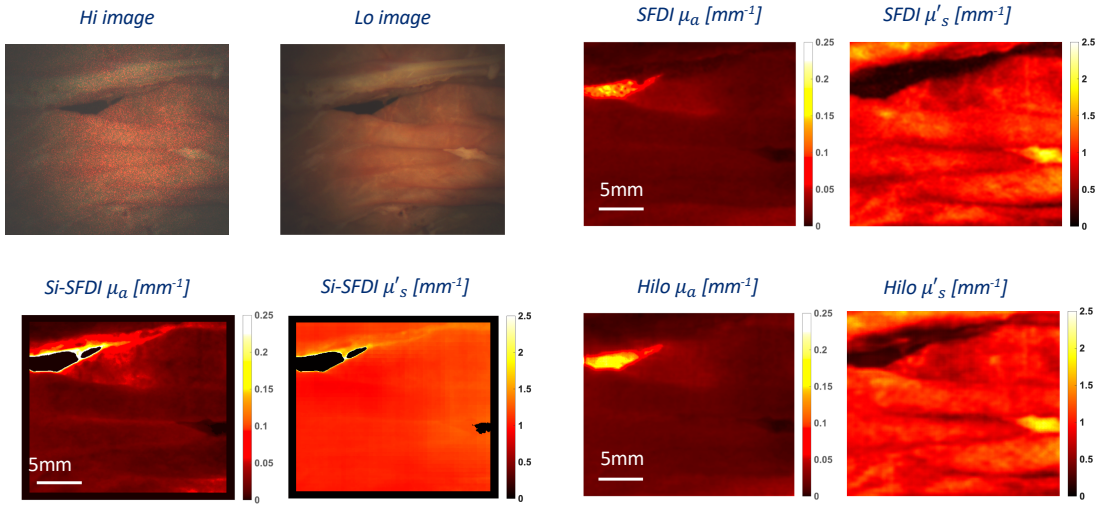

Figure 5-6. Top left: example Hi and Lo image taken by the left camera in the stereo pair. Top right: corresponding ground truth optical property maps using conventional SFDI. Bottom left: results from si-SFDI processing. Bottom right: results using HiLo illumination and machine learning. Note that the optical property maps are smaller than the original images due to the field-of-view mismatch between left and right cameras, which is cropped during depth reconstruction.

We also study the effects of HiLo illumination compared to Hi and Lo alone. The errors in terms of NMAE are shown in Table 5-I for all the images in the testing set. We observe an accuracy improvement in both μ'_s and μ_a for the proposed HiLo illumination.

In terms of acquisition and processing speed, the proposed set up acquires a single

Table 5-I. Normalized mean absolute error (NMAE) in absorption and reduced scattering coefficients using different illumination schemes, including HiLo (proposed), Lo only, and Hi only. HiLo outperforms the other benchmarks for both absorption and scattering. Best results are shown in bold.

Input	μ_a *error*	μ'_s *error*
HiLo	**0.1595**	**0.1608**
Lo only	0.1721	0.1731
Hi only	0.1889	0.1905

image pair at an average exposure time of 50ms at a 5cm working distance. The neural network takes approximately 8ms to process one image on a single GPU, and the ultrafast lookup table takes 20ms to predict optical properties from reflectance.

5.3.3 Discussion

In this section, we describe the hardware setup and processing flow for fast and accurate endoscopic optical property mapping enabled by machine learning and HiLo illumination. The proposed method uses two images - a DC (Lo) and a AC (Hi) image - to predict the optical properties of tissues. Compared to existing techniques for endoscopic optical property mapping using structured light [58], the proposed setup requires minimal modifications to the commercially available da Vinci endoscope using fiber-coupled, coherent illumination. It simultaneously obtains depth information using stereo vision, which can be applied to reflectance measurements for profilometry correction. Moreover, with the help of machine learning, the HiLo endoscope improves upon the performance of model-based si-SFDI in terms of speed, image quality, and accuracy. With further optimization in illumination power and coupling, the proposed technique has the potential to achieve real-time acquisition and processing.

Currently, passive stereo using unstructured illumination is employed for depth estimation. Studies have demonstrated the feasibility of using random speckle patterns generated by a coherent source and a diffusing object for active stereo [137, 138]. Unfortunately, we do not observe any correspondences of speckle patterns between left and right images using the current setup (Fig. 5-7(a)). We hypothesize that this

is because subjective speckles are much bigger than objective speckles in our scenario. Objective speckles are formed when a rough surface is illuminated by a coherent light source, while subjective speckles are formed by the resolution elements of a lens. When using laser speckles as structured illumination for active stereo or si-SFDI, we refer to objective speckles because they form at the object plane and do not depend on the imaging system. In contrast, subjective speckles form at the image plane, which is a statistical process that is related to the lens and the imaging setup [123]. Both types of speckles are present in camera images of a projected speckle pattern, and if not properly removed, subjective speckles could be detrimental to applications that rely on imaging of objective speckles [139]. Compared to the original si-SFDI [53] and other published active stereo systems [137], the setup in this study has a much shorter working distance and smaller magnification and aperture size, which are contributing factors to the relative dimensions of objective and subjective speckles. The average size of objective speckles can be expressed as:

$$d_o = \frac{\lambda z}{D}, \tag{5.2}$$

where λ is the wavelength, z is the observation distance (or screen) from the diffusing surface, and D is the spot size of the laser beam on the surface. The average size of subjective speckles is:

$$d_s = (1+m) \cdot \frac{\lambda}{2NA} \approx (1+m) \cdot F_{\#} \cdot \lambda, \tag{5.3}$$

where NA and $F_{\#}$ stand for the numerical aperture and the f-number of the system, respectively. The endoscope used in this study produces large subjective speckles due to the small aperture size (or large f-number) and large magnification. The average size of the subjective speckles in this case is measured to be approximately 0.120mm, or three image pixels.

To measure the size of objective speckles, one can place a camera sensor directly at the object plane, which is not easily implementable in our case. However, according to Eq. 5.2, we can simply alter the laser spot size on the illumination port of the endoscope to change the objective speckle size. When focusing the laser beam onto the incident surface of a fiber bundle (Gulf Photonics), which is then butt coupled to the endoscope, we notice a significant increase in the objective speckle size and correspondences between left and right images (Fig. 5-7(b)). The illumination spot size is approximately 0.5mm in diameter. When mounted externally on the same level as the tip of the endoscope, the fiber bundle with a 0.5mm diameter spot size produces large speckles at a even higher contrast (Fig. 5-7(c)). Therefore, a next step of the study is to use a similar setup to Fig. 5-7(c) and explore its potential for active stereo.

In Fig. 5-6, compared to conventional SFDI as ground truth, HiLo endoscope with machine learning produces optical properties that are more accurate than si-SFDI. However, there are noticeable discrepancies between HiLo and SFDI in some regions of the tissue. It could be caused by the different illumination angles between lasers (0 degrees) and the projector (25 degrees). It could also be attributed to the wavelength mismatch. The red LED diode in the projector has a center wavelength of 617nm and

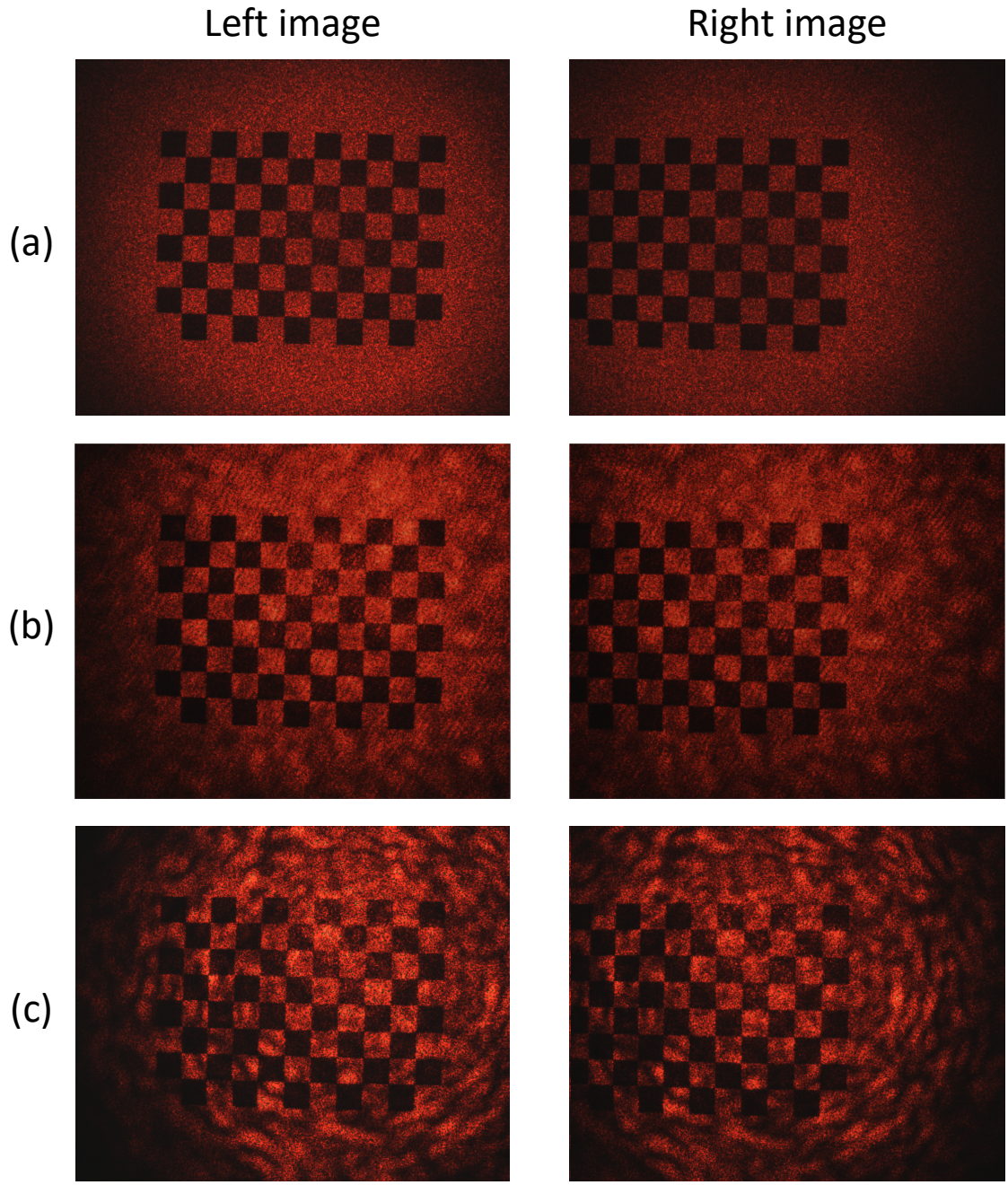

Figure 5-7. Acquired speckle patterns using (a) 1mm diameter multimode fiber butt-coupled to endoscope's rigid light guide; (b) 1mm diameter fiber bundle butt-coupled to endoscope, illumination spot size = 0.5mm; (c) 1mm diameter fiber bundle mounted externally at the same level as the tip of the endoscope, spot size = 0.5mm. Images of the same checkerboard pattern captured by the left camera and the right camera are shown in the left and right column, respectively. No correspondences in the pattern are observed in (a). Correspondences can be seen in (b) and (c), with the latter producing a higher contrast in the objective speckle pattern.

a bandwidth of 18nm, while the laser diode is 637nm.

Furthermore, in Table 5-I, we see that proposed HiLo illumination improves the accuracy of using Hi and Lo alone. However, the difference is not as big as expected, possibly due to the small size and/or low contrast of objective speckles relative to subjective speckles. It is important to reiterate the fact that si-SFDI measures optical properties by characterizing tissue response to *objective* speckles, which is independent of the imaging setup. Due to the small size (approximately 1 pixel) relative to objective speckles (3 to 5 pixels), subjective speckles are reduced by median filtering in [53]. Similarly, we hypothesize that larger objective speckles could be useful for more accurate optical property mapping using HiLo illumination.

In addition, due to hardware limitations, the current setup only measures the optical properties at 636/637nm wavelength. However, future experiments could incorporate green and blue lasers with long coherence lengths to enable simultaneous RGB speckle illumination. Because color cameras and narrow bandwidth lasers are used, tissue response to each individual wavelength could be isolated by separating the three color channels of a single Hi image.

To train a more robust machine learning model, we could include more samples or a larger variety of tissue types. A more sophisticated algorithm could be employed to accurately model the image translation process from HiLo images to optical property maps [140]. Depth estimation could also be improved with a state-of-the-art deep learning technique [141, 142]. Moreover, the proposed method is a multi-step process that estimates depth, tissue response, and optical property separately (Fig. 5-4).

An end-to-end model could be trained to directly predict profile-corrected optical properties from left and right HiLo pairs. This would improve the processing speed while inevitably requiring more training data. It would also come at the cost of model interpretability because it would bypass all the intermediate steps, making it impossible to isolate the source of potential errors.

Summary and Conclusions

5.4 Aim 1: Single-shot Tissue Optical Property and Biomarker Measurements Enabled by Machine Learning

In Chapter 2, we presented GANPOP, which was the first study to estimate optical properties directly from a single structured illumination image using a data-driven model. We demonstrated advantages of AC versus DC illumination to determine optical properties via an adversarial learning approach, and we benchmarked our technique against various model-based and learning-based methods. Moreover, we acquired and made publicly-available a dataset of registered conventional wide-field images, structured-illumination images, and ground-truth optical properties of a variety of *ex vivo* and *in vivo* tissues.

In Chapter 3, we extended the adversarial deep learning approach to predicting wide-field tissue oxygenation. The proposed method, called OxyGAN, estimated tissue oxygenation from single structured light images at two near-infrared wavelengths

in an end-to-end fashion. It demonstrated superior accuracy and image quality to various benchmarks, including model-based SSOP and a multi-step process that estimates optical properties with two separate GANPOP networks followed by Beer-Lambert fitting. OxyGAN correctly detected changes in oxygenation over time and was relatively insensitive to motion. We additionally achieved real-time processing by optimizing trained OxyGAN models for reduced latency and fast inference.

The techniques presented in Chapters 2 and 3 aimed to tackle the first challenge that limits the clinical utility of SFDI, which was the trade-off between acquisition speed and accuracy. We have achieved highly accurate biomarker mapping from single frames in real time using a relatively small dataset. The proposed methods also have the potential to reduce the cost of SFDI systems. For example, harnessing the power of GANPOP and OxyGAN, future systems could replace projectors with LEDs and a sinusoidal mask for low-cost, single-pattern projection.

5.5 Aim 2: Endoscopic Imaging of Tissue Biomarkers Using Speckle Illumination Spatial Frequency Domain Imaging

In Chapter 4, a signal processing technique, called si-SFDI, was presented to accurately estimate optical properties from a series of random laser speckle patterns. We conducted comprehensive analyses of the effects of surface topography, blood flow, and the number of input patterns on the performance of the method. The method also

demonstrated stability over time and a spatial resolution similar to that of conventional SFDI.

In contrast to previously published research [121], si-SFDI can unambiguously decouple absorption and scattering by relating information derived from power spectral density to Monte-Carlo simulated spatial frequency response. More importantly, it is capable of accurate measurements even when applied to heterogeneous specimens. Additionally, si-SFDI is a projector-free technique that uses coherent light sources to generate speckle patterns as structured illumination, making it more amenable to space-constrained clinical applications.

In Chapter 5, we first showed initial results of si-SDFI incorporated into a clinical colonoscope and discussed current limitations. We then developed a novel computational imaging technology that allows tissue optical properties to be rapidly imaged through a stereo endoscope with fiber-delivered speckle illumination and machine learning. The proposed HiLo SFDI technology had several attributes that would make it a practical solution for endoscopic optical property mapping: (1) it required minimal and low-cost modifications to clinically available systems–in some cases just the addition of a laser illumination source would be sufficient, (2) it was capable of rapid data acquisition and processing via two snapshots by using deep learning algorithms, (3) it was capable of producing high-quality profile-corrected optical property maps by combining speckle illumination and depth estimation from stereo vision.

5.6 Limitations and Future Work

Throughout this thesis, we aim to resolve the major issues associated with conventional SFDI that prevent it from being widely adopted in clinical settings. Specifically, we have developed a novel data-driven algorithm for real-time, high-quality imaging of tissue biomarkers. We have also demonstrated the feasibility of using compact coherent light sources and random speckle patterns for quantitative endoscopic imaging of tissues.

Despite their merits, the techniques described in this thesis do not come without limitations. For example, GANPOP and OxyGAN are sensitive to noise and ambient lighting. One major benefit of three-phase demodulation is to remove the effects of dark current and background light, which is not accounted for in single-snapshot techniques. Second, all the training and testing images were acquired using a commercial SFDI system, and thus we anticipate the trained models to be sensitive to changes in imaging setups, such as image resolution, field-of-view, and projected spatial frequencies. Moreover, OxyGAN is sensitive to rapid motion due to the limited frame rate of the camera. Because sequential flashing is used to switch between the two wavelengths, any large displacements between frames would introduce errors to its estimations.

The major issue of si-SFDI is the slow imaging and processing speed. It currently uses 4 input images to produce accurate results; the average processing time is 2 seconds per image due to the computationally expensive, window-based ACF estimations. Another drawback is the poor image quality in regions with high levels of heterogeneity and complex topography. Furthermore, it is important to note that the si-SFDI imaging

geometry described in Chapter 4 differs greatly from a conventional endoscope in terms of magnification, working distance, and numerical aperture. Thus, changes in the relative size and contrast between subjective and objective speckles could become a source of error. Although the HiLo endoscope described in Chapter 5 attempts to alleviate these limitations with stereo vision and machine learning, additional work is needed to corroborate the system design and enhance its performance.

In the future, for GANPOP and OxyGAN, a more generalizable model that would work on a range of imaging systems could be trained using domain adaptation techniques. A wider range of optical properties or tissue types could also be incorporated into the training set. A larger dataset would also enable us to train on full-size images instead of patches, which would reduce stitching artifacts at the patch boundaries. To reduce the effects of motion artifacts in OxyGAN, tissues could be illuminated with 660nm and 850nm wavelength simultaneously and imaged with two cameras and different spectral filters. Moreover, in our studies, we examined the potential of GANPOP and OxyGAN to predict profile-corrected biomarker maps without explicitly outputting sample profilometry. One future direction is to use machine learning for high-quality shape measurements from single-shot structured light images [143]. We would need to alter the direction of the projected sinusoid or design a different pattern that is sensitive to changes in topography [144, 48]. Furthermore, large-scale studies could be conducted to investigate the diagnostic quality of these methods. One potential application is diabetic foot ulcer staging. In this scenario, our proposed techniques would be assessed based on the ability to produce accurate contrast between normal

and ulcer tissues and the sensitivity to patient melanin levels, age, and weight.

Discussed in more details in Chapter 5, additional research is needed to improve the performance of the HiLo illumination endoscope. We will assess the utility of larger objective speckles for both depth and optical property reconstructions. To produce high objective speckle contrast, we will externally mount the fiber bundle at the same level as the tip of the endoscope. The size of the bundle (1mm) is small compared to the endoscope (12mm diameter). We anticipate that objective speckles with a similar size and contrast could be achieved in future prototypes by either inserting the bundle through an instrument port or replacing the rigid endoscope light guide to prevent loss of coherence through butt-coupling. The difference in illumination angles between lasers and the projector could be reduced by projecting patterns through a second laparoscope [145]. A more sophisticated training scheme could be employed to account for out-of-distribution regions caused by this angle mismatch [140]. Furthermore, the setup could be evaluated during a pig anastomosis surgery and *in vivo* human studies.

In addition to compact size, fiber-delivered laser speckles are advantageous to projector-based sinusoidal patterns because they can sample a wide range of spatial frequencies and produce high spatial frequencies at a large depth of field. So far, si-SFDI only uses a narrow spectral band for optical property estimations. An interesting area of research would be to explore random laser speckles for extracting sub-diffusive, depth-resolved information with high spatial frequency content [146, 147].

To improve the clinical utility of SFDI, we have been studying the use of high spatial frequency structured illumination for slide-free, wide-field imaging of tissue

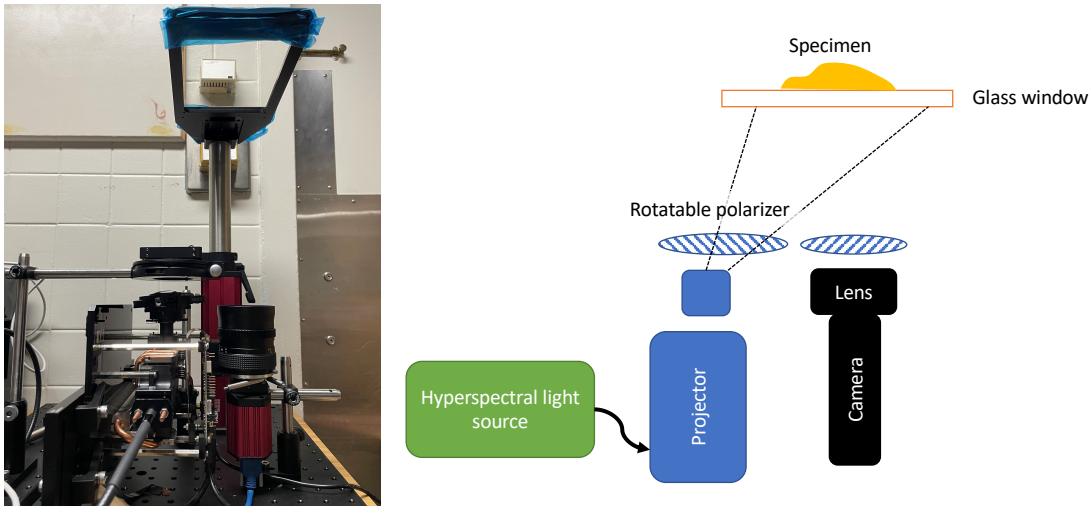

Figure 5-8. Inverted system for hyperspectral, structured light, and polarized light imaging. A glass window is used for tissue flattening and specularity removal.

histology. Recent studies have mostly focused on the diagnostic quality of high spatial frequency, sub-diffusive SFDI applied to breast cancer tissues [37, 36]. We are interested in exploring its potential for gastrointestinal tissues. For this purpose, we have built an inverted system that is capable of hyperspectral, high spatial frequency SFDI, and polarized light image (Fig. 5-8). To correlate SFDI signals with pathology, we will perform grid biopsy on specimens and use white-light microscopy images of H&E-stained slides as ground truth. The target output could be either tissue types or histological primitives, such as nuclei density or nuclear-cytoplasmic ratio.

In this thesis, we have demonstrated the utility of structured light combined with data-driven methods for quantitative tissue imaging. Specifically, we focused on tackling the technical challenges of Spatial Frequency Domain Imaging, which is a powerful optical imaging tool for tissue biomarker measurements. Although SFDI as a medical imaging modality is still in early stages of development, we hope the

approaches described in this work help to facilitate the clinical translation of SFDI and the future research in this area.

References

[1] A. Mazhar, S. Dell, D. J. Cuccia, S. Gioux, A. J. Durkin, J. V. Frangioni, and B. J. Tromberg, "Wavelength optimization for rapid chromophore mapping using spatial frequency domain imaging," *Journal of Biomedical Optics*, vol. 15, no. 6, p. 061716, Dec. 2010.

[2] D. J. Cuccia, F. P. Bevilacqua, A. J. Durkin, F. R. Ayers, and B. J. Tromberg, "Quantitation and mapping of tissue optical properties using modulated imaging," *Journal of Biomedical Optics*, vol. 14, no. 2, p. 024012, 2009.

[3] M. T. Chen, F. Mahmood, J. A. Sweer, and N. J. Durr, "Ganpop: Generative adversarial network prediction of optical properties from single snapshot wide-field images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1988–1999, 2020.

[4] T. D. O'Sullivan, A. E. Cerussi, B. J. Tromberg, and D. J. Cuccia, "Diffuse optical imaging using spatially and temporally modulated light," *Journal of biomedical optics*, vol. 17, no. 7, p. 071311, 2012.

[5] S. S. Streeter, B. W. Maloney, R. A. Zuurbier, W. A. Wells, R. J. Barth, K. D. Paulsen, and B. W. Pogue, "Optical scatter imaging of resected breast tumor

structures matches the patterns of micro-computed tomography," *Physics in Medicine & Biology*, vol. 66, no. 11, p. 115021, 2021.

[6] G. A. Murphy, R. P. Singh-Moon, A. Mazhar, D. J. Cuccia, V. L. Rowe, and D. G. Armstrong, "Quantifying dermal microcirculatory changes of neuropathic and neuroischemic diabetic foot ulcers using spatial frequency domain imaging: a shade of things to come?" *BMJ Open Diabetes Research and Care*, vol. 8, no. 2, p. e001815, 2020.

[7] S. Gioux, A. Mazhar, B. T. Lee, S. J. Lin, A. M. Tobias, D. J. Cuccia, A. Stockdale, R. Oketokoun, Y. Ashitate, E. Kelly, M. Weinmann, N. J. Durr, L. A. Moffitt, A. J. Durkin, B. J. Tromberg, and J. V. Frangioni, "First-in-human pilot study of a spatial frequency domain oxygenation imaging system," *Journal of Biomedical Optics*, vol. 16, no. 8, 2011.

[8] G. Keiser, *Light-Tissue Interactions*. Singapore: Springer Singapore, 2016, pp. 147–196. [Online]. Available: https://doi.org/10.1007/978-981-10-0945-7_6

[9] "Ece 532." [Online]. Available: https://omlc.org/classroom/ece532/

[10] A. M. Laughney, V. Krishnaswamy, E. J. Rizzo, M. C. Schwab, R. J. Barth, D. J. Cuccia, B. J. Tromberg, K. D. Paulsen, B. W. Pogue, and W. A. Wells, "Spectral discrimination of breast pathologies in situusing spatial frequency domain imaging," *Breast Cancer Research*, vol. 15, no. 4, p. R61, 2013.

[11] P. Bönöczk, G. Panczel, and Z. Nagy, "Vinpocetine increases cerebral blood flow and oxygenation in stroke patients: a near infrared spectroscopy and transcranial doppler study," *European journal of ultrasound*, vol. 15, no. 1-2, pp. 85–91, 2002.

[12] S. L. Jacques, "Optical properties of biological tissues: a review," *Physics in Medicine & Biology*, vol. 58, no. 11, p. R37, 2013.

[13] R. Richards-Kortum and E. Sevick-Muraca, "Quantitative optical spectroscopy for tissue diagnosis," *Annual Review of Physical Chemistry*, vol. 47, no. 1, pp. 555–606, 1996.

[14] R. A. Drezek, M. Guillaud, T. G. Collier, I. Boiko, A. Malpica, C. E. MacAulay, M. Follen, and R. R. Richards-Kortum, "Light scattering from cervical cells throughout neoplastic progression: influence of nuclear morphology, dna content, and chromatin texture," *Journal of biomedical optics*, vol. 8, no. 1, pp. 7–17, 2003.

[15] B. W. Maloney, D. M. McClatchy, B. W. Pogue, K. D. Paulsen, W. A. Wells, and R. J. Barth, "Review of methods for intraoperative margin detection for breast conserving surgery," *Journal of biomedical optics*, vol. 23, no. 10, p. 100901, 2018.

[16] J. R. Mourant, M. Canpolat, C. Brocker, O. Esponda-Ramos, T. M. Johnson, A. Matanock, K. Stetter, and J. P. Freyer, "Light scattering from cells: the contribution of the nucleus and the effects of proliferative status," *Journal of biomedical optics*, vol. 5, no. 2, pp. 131–138, 2000.

[17] Z. A. Steelman, D. S. Ho, K. K. Chu, and A. Wax, "Light-scattering methods for tissue diagnosis," *Optica*, vol. 6, no. 4, pp. 479–489, Apr 2019.

[18] M. Siegemund, J. Van Bommel, and C. Ince, "Assessment of regional tissue oxygenation," *Intensive care medicine*, vol. 25, no. 10, p. 1044, 1999.

[19] A. J. Lin, M. A. Koike, K. N. Green, J. G. Kim, A. Mazhar, T. B. Rice, F. M. LaFerla, and B. J. Tromberg, "Spatial frequency domain imaging of intrinsic optical property contrast in a mouse model of alzheimer's disease," *Annals of Biomedical Engineering*, vol. 39, no. 4, pp. 1349–1357, Apr 2011.

[20] N. Shah, A. Cerussi, C. Eker, J. Espinoza, J. Butler, J. Fishkin, R. Hornung, and B. Tromberg, "Noninvasive functional optical spectroscopy of human breast tissue," *Proceedings of the National Academy of Sciences*, vol. 98, no. 8, pp. 4420–4425, 2001.

[21] D. J. Cuccia, F. Bevilacqua, A. J. Durkin, F. R. Ayers, and B. J. Tromberg, "Quantitation and mapping of tissue optical properties using modulated imaging," *Journal of Biomedical Optics*, vol. 14, no. 2, p. 024012, Apr. 2009.

[22] M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties," *Applied optics*, vol. 28, no. 12, pp. 2331–2336, 1989.

[23] T. J. Farrell, M. S. Patterson, and B. Wilson, "A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo," *Medical physics*, vol. 19, no. 4, pp. 879–888, 1992.

[24] N. Dögnitz and G. Wagnières, "Determination of tissue optical properties by steady-state spatial frequency-domain reflectometry," *Lasers in medical science*, vol. 13, no. 1, pp. 55–65, 1998.

[25] S. Gioux, A. Mazhar, and D. J. Cuccia, "Spatial frequency domain imaging in

2019: principles, applications, and perspectives," *Journal of biomedical optics*, vol. 24, no. 7, p. 071613, 2019.

[26] A. Bassi, D. J. Cuccia, A. J. Durkin, and B. J. Tromberg, "Spatial shift of spatially modulated light projected on turbid media," *JOSA A*, vol. 25, no. 11, pp. 2833–2839, 2008.

[27] S. L. Jacques and L. Wang, "Monte carlo modeling of light transport in tissues," in *Optical-thermal response of laser-irradiated tissue*. Springer, 1995, pp. 73–100.

[28] A. Kienle and M. S. Patterson, "Determination of the optical properties of turbid media from a single monte carlo simulation," *Physics in Medicine & Biology*, vol. 41, no. 10, p. 2221, 1996.

[29] V. Pera, K. Karrobi, S. Tabassum, F. Teng, and D. Roblyer, "Optical property uncertainty estimates for spatial frequency domain imaging," *Biomedical Optics Express*, vol. 9, no. 2, pp. 661–678, 2018.

[30] S. Gioux, A. Mazhar, D. J. Cuccia, A. J. Durkin, B. J. Tromberg, and J. V. Frangioni, "Three-Dimensional Surface Profile Intensity Correction for Spatially-Modulated Imaging," *Journal of biomedical optics*, vol. 14, no. 3, p. 034045, 2009.

[31] G. Indebetouw, "Profile measurement using projection of running fringes," *Applied Optics*, vol. 17, no. 18, pp. 2930–2933, 1978.

[32] M. Takeda and K. Mutoh, "Fourier transform profilometry for the automatic measurement of 3-d object shapes," *Applied optics*, vol. 22, no. 24, pp. 3977–3982, 1983.

[33] V. Srinivasan, H.-C. Liu, and M. Halioua, "Automated phase-measuring profilometry of 3-d diffuse objects," *Applied optics*, vol. 23, no. 18, pp. 3105–3108, 1984.

[34] Y. Zhao, S. Tabassum, S. Piracha, M. S. Nandhu, M. Viapiano, and D. Roblyer, "Angle correction for small animal tumor imaging with spatial frequency domain imaging (sfdi)," *Biomedical optics express*, vol. 7, no. 6, pp. 2373–2384, 2016.

[35] C. K. Hayakawa, K. Karrobi, V. E. Pera, D. M. Roblyer, and V. Venugopalan, "Optical sampling depth in the spatial frequency domain," *Journal of biomedical optics*, vol. 24, no. 7, p. 071603, 2018.

[36] D. M. McClatchy, E. J. Rizzo, W. A. Wells, P. P. Cheney, J. C. Hwang, K. D. Paulsen, B. W. Pogue, and S. C. Kanick, "Wide-field quantitative imaging of tissue microstructure using sub-diffuse spatial frequency domain imaging," *Optica*, vol. 3, no. 6, pp. 613–621, 2016.

[37] S. S. Streeter, B. Hunt, R. A. Zuurbier, W. A. Wells, K. D. Paulsen, and B. W. Pogue, "Developing diagnostic assessment of breast lumpectomy tissues using radiomic and optical signatures," *Scientific reports*, vol. 11, no. 1, pp. 1–10, 2021.

[38] J. Vervandier and S. Gioux, "Single snapshot imaging of optical properties," *Biomedical Optics Express*, vol. 4, no. 12, pp. 2938–2944, Nov. 2013.

[39] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[40] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang,

G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[41] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Neural networks for perception*. Elsevier, 1992, pp. 65–93.

[42] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, 2016.

[43] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *towards data science*, vol. 6, no. 12, pp. 310–316, 2017.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, p. 234–241, 2015.

[45] S. Panigrahi and S. Gioux, "Machine learning approach for rapid and accurate estimation of optical properties using spatial frequency domain imaging," *Journal of Biomedical Optics*, vol. 24, no. 7, p. 071606, 2018.

[46] Y. Zhao, Y. Deng, F. Bao, H. Peterson, R. Istfan, and D. Roblyer, "Deep learning model for ultrafast multifrequency optical property extractions for spatial frequency domain imaging," *Optics letters*, vol. 43, no. 22, pp. 5669–5672, 2018.

[47] Y. Zhao, Y. Deng, S. Yue, M. Wang, B. Song, and Y. Fan, "Direct mapping from diffuse reflectance to chromophore concentrations in multi-fx spatial frequency domain imaging (sfdi) with a deep residual network (drn)," *Biomedical Optics Express*, vol. 12, no. 1, pp. 433–443, 2021.

[48] E. Aguénounon, J. T. Smith, M. Al-Taher, M. Diana, X. Intes, and S. Gioux, "Real-time, wide-field and high-quality single snapshot imaging of optical properties with profile correction using deep learning," *Biomedical Optics Express*, vol. 11, no. 10, pp. 5701–5716, 2020.

[49] M. T. Chen and N. J. Durr, "Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning," *Journal of Biomedical Optics*, vol. 25, no. 11, p. 112907, 2020.

[50] M. T. Chen and N. J. Durr, "Quantitative tissue property measurements with structured illumination and deep learning," in *2020 IEEE Photonics Conference (IPC)*. IEEE, pp. 1–2.

[51] M. T. Chen and N. J. Durr, "Real-time oxygenation mapping from structured light imaging with deep learning," in *Optical Tomography and Spectroscopy of Tissue XIV*, vol. 11639. International Society for Optics and Photonics, 2021, p. 116391D.

[52] M. T. Chen, T. L. Bobrow, and N. J. Durr, "Towards sfdi endoscopy with structured illumination from randomized speckle patterns," in *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XIX*, vol. 11631. International Society for Optics and Photonics, 2021, p. 116310Y.

[53] M. T. Chen, M. Papadakis, and N. J. Durr, "Speckle illumination sfdi for projector-free optical property mapping," *Optics letters*, vol. 46, no. 3, pp. 673–676, 2021.

[54] M. R. Pharaon, T. Scholz, S. Bogdanoff, D. Cuccia, A. J. Durkin, D. B. Hoyt,

and G. R. D. Evans, "Early detection of complete vascular occlusion in a pedicle flap model using quantitative [corrected] spectral imaging," *Plastic and Reconstructive Surgery*, vol. 126, no. 6, pp. 1924–1935, Dec. 2010.

[55] C. Weinkauf, A. Mazhar, K. Vaishnav, A. A. Hamadani, D. J. Cuccia, and D. G. Armstrong, "Near-instant noninvasive optical imaging of tissue perfusion for vascular assessment," *Journal of Vascular Surgery*, vol. 69, no. 2, pp. 555 – 562, 2019.

[56] M. Kaiser, A. Yafi, M. Cinat, B. Choi, and A. J. Durkin, "Noninvasive assessment of burn wound severity using optical technology: a review of current and future modalities," *Burns: Journal of the International Society for Burn Injuries*, vol. 37, no. 3, pp. 377–386, May 2011.

[57] A. Yafi, F. K. Muakkassa, T. Pasupneti, J. Fulton, D. J. Cuccia, A. Mazhar, K. N. Blasiole, and E. N. Mostow, "Quantitative skin assessment using spatial frequency domain imaging (sfdi) in patients with or at high risk for pressure ulcers," *Lasers in Surgery and Medicine*, vol. 49, no. 9, pp. 827–834, 2017.

[58] J. P. Angelo, M. van de Giessen, and S. Gioux, "Real-time endoscopic optical properties imaging," *Biomedical Optics Express*, vol. 8, no. 11, pp. 5113–5126, Oct. 2017.

[59] S. Nandy, W. Chapman, R. Rais, I. González, D. Chatterjee, M. Mutch, and Q. Zhu, "Label-free quantitative optical assessment of human colon tissue using spatial frequency domain imaging," *Techniques in Coloproctology*, vol. 22, no. 8, pp. 617–621, Aug. 2018.

[60] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3, pp. 257–273, Sep. 2017.

[61] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[62] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.

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

[64] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[65] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.

[66] A. Diaz-Pinto, A. Colomer, V. Naranjo, S. Morales, Y. Xu, and A. F. Frangi, "Retinal image synthesis and semi-supervised learning for glaucoma assessment,"

IEEE Transactions on Medical Imaging, pp. 1–1, 2019.

[67] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, June 2018.

[68] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, June 2018.

[69] J. Swartling, J. S. Dam, and S. Andersson-Engels, "Comparison of spatially and temporally resolved diffuse-reflectance measurement systems for determination of biomedical optical properties," *Appl. Opt.*, vol. 42, no. 22, pp. 4612–4620, Aug 2003.

[70] J. R. Weber, D. J. Cuccia, A. J. Durkin, and B. J. Tromberg, "Noncontact imaging of absorption and scattering in layered tissue using spatially modulated structured light," *Journal of Applied Physics*, vol. 105, no. 10, p. 102028, 2009.

[71] G. M. Palmer, R. J. Viola, T. Schroeder, P. S. Yarmolenko, M. W. Dewhirst, and N. Ramanujam, "Quantitative Diffuse Reflectance and Fluorescence Spectroscopy: A Tool to Monitor Tumor Physiology In Vivo," *Journal of biomedical optics*, vol. 14, no. 2, p. 024010, 2009.

[72] G. Jones, N. T. Clancy, Y. Helo, S. Arridge, D. S. Elson, and D. Stoyanov, "Bayesian estimation of intrinsic tissue oxygenation and perfusion from rgb

images," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1491–1501, July 2017.

[73] M. van de Giessen, J. P. Angelo, and S. Gioux, "Real-time, profile-corrected single snapshot imaging of optical properties," *Biomedical Optics Express*, vol. 6, no. 10, pp. 4051–4062, Sep. 2015.

[74] S. G. Swapnesh Panigrahi, "Machine learning approach for rapid and accurate estimation of optical properties using spatial frequency domain imaging," *Journal of Biomedical Optics*, vol. 24, no. 7, pp. 1 – 6 – 6, 2018.

[75] M. Martinelli, A. Gardner, D. Cuccia, C. Hayakawa, J. Spanier, and V. Venugopalan, "Analysis of single Monte Carlo methods for prediction of reflectance from turbid media," *Optics Express*, vol. 19, no. 20, pp. 19 627–19 642, Sep. 2011.

[76] R. Chen, F. Mahmood, A. Yuille, and N. J. Durr, "Rethinking monocular depth estimation with adversarial training," *arXiv preprint arXiv:1808.07528*, 2018.

[77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[78] T. M. Quan, D. G. C. Hildebrand, and W.-K. Jeong, "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics," *CoRR*, vol. abs/1612.05360, 2016.

[79] T. Miyato and M. Koyama, "cgans with projection discriminator," *arXiv preprint arXiv:1802.05637*, 2018.

[80] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[81] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[82] J. A. Sweer, M. T. Chen, K. J. Salimian, R. J. Battafarano, and N. J. Durr, "Wide-field optical property mapping and structured light imaging of the esophagus with spatial frequency domain imaging," *Journal of biophotonics*, p. e201900005, 2019.

[83] F. Ayers, A. Grant, D. Kuo, D. J. Cuccia, and A. J. Durkin, "Fabrication and characterization of silicone-based tissue phantoms with tunable optical properties in the visible and near infrared domain," in *Design and Performance Validation of Phantoms Used in Conjunction with Optical Measurements of Tissue*, vol. 6870. International Society for Optics and Photonics, Feb. 2008, p. 687007.

[84] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, "Are gans created equal? a large-scale study," in *Advances in neural information processing systems*, 2018, pp. 700–709.

[85] F. Mahmood, R. Chen, and N. J. Durr, "Unsupervised reverse domain adaptation for synthetic medical images via adversarial training," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2572–2581, 2018.

[86] F. Mahmood, R. Chen, S. Sudarsky, D. Yu, and N. J. Durr, "Deep learning with cinematic rendering: fine-tuning deep neural networks using photorealistic

medical images," *Physics in Medicine & Biology*, vol. 63, no. 18, p. 185012, sep 2018.

[87] H. L. Edmonds Jr, B. L. Ganzel, and E. H. Austin, "Cerebral oximetry for cardiac and vascular surgery," in *Seminars in Cardiothoracic and Vascular Anesthesia*, vol. 8, no. 2. SAGE Publications Sage CA: Los Angeles, CA, 2004, pp. 147–166.

[88] H. M. Swartz and J. F. Dunn, "Measurements of oxygen in tissues: overview and perspectives on methods," in *Oxygen Transport to Tissue XXIV*. Springer, 2003, pp. 1–12.

[89] M. Sair, P. J. Etherington, C. P. Winlove, and T. W. Evans, "Tissue oxygenation and perfusion in patients with systemic sepsis," *Critical Care Medicine*, vol. 29, no. 7, pp. 1343–1349, 2001.

[90] J. Ditzel and E. Standl, "The problem of tissue oxygenation in diabetes mellitus: Its relation to the early functional changes in the microcirculation of diabetic subjects," *Acta Medica Scandinavica*, vol. 197, no. S578, pp. 49–58, 1975.

[91] G. Pitsiou, G. Kyriazis, O. Hatzizisi, P. Argyropoulou, E. Mavrofridis, and D. Patakas, "Tumor necrosis factor—alpha serum levels, weight loss and tissue oxygenation in chronic obstructive pulmonary disease," *Respiratory Medicine*, vol. 96, no. 8, pp. 594–598, 2002.

[92] Y. Mendelson, "Pulse oximetry: theory and applications for noninvasive monitoring," *Clinical Chemistry*, vol. 38, no. 9, pp. 1601–1607, 1992.

[93] S. Gioux, A. Stockdale, R. Oketokoun, Y. Ashitate, N. J. Durr, L. A. Moffitt,

J. V. Frangioni, A. Mazhar, B. J. Tromberg, A. J. Durkin *et al.*, "First-in-human pilot study of a spatial frequency domain oxygenation imaging system," *Journal of Biomedical Optics*, vol. 16, no. 8, p. 086015, 2011.

[94] H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise, *Near-infrared spectroscopy: principles, instruments, applications*. John Wiley & Sons, 2008.

[95] D. A. Benaron, I. H. Parachikov, S. Friedland, R. Soetikno, J. Brock-Utne, P. J. Van Der Starre, C. Nezhat, M. K. Terris, P. G. Maxim, J. J. Carson *et al.*, "Continuous, noninvasive, and localized microvascular tissue oximetry using visible light spectroscopy," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 100, no. 6, pp. 1469–1475, 2004.

[96] J. M. Murkin and M. Arango, "Near-infrared spectroscopy as an index of brain and tissue oxygenation," *British Journal of Anaesthesia*, vol. 103, no. suppl_1, pp. i3–i13, 2009.

[97] J. T. Nguyen, S. J. Lin, A. M. Tobias, S. Gioux, A. Mazhar, D. J. Cuccia, Y. Ashitate, A. Stockdale, R. Oketokoun, N. J. Durr *et al.*, "A novel pilot study using spatial frequency domain imaging to assess oxygenation of perforator flaps during reconstructive breast surgery," *Annals of Plastic Surgery*, vol. 71, no. 3, p. 308, 2013.

[98] K. J. Zuzak, M. D. Schaeberle, E. N. Lewis, and I. W. Levin, "Visible reflectance hyperspectral imaging: characterization of a noninvasive, in vivo system for determining tissue perfusion," *Analytical Chemistry*, vol. 74, no. 9, pp. 2021–2028, 2002.

[99] S. Srinivasan, B. W. Pogue, S. Jiang, H. Dehghani, and K. D. Paulsen, "Spectrally constrained chromophore and scattering near-infrared tomography provides quantitative and robust reconstruction," *Applied optics*, vol. 44, no. 10, pp. 1858–1869, 2005.

[100] J. Q. M. Nguyen, C. Crouzet, T. Mai, K. Riola, D. Uchitel, L.-H. L. Liaw, N. Bernal, A. Ponticorvo, B. Choi, and A. J. Durkin, "Spatial frequency domain imaging of burn wounds in a preclinical model of graded burn severity," *Journal of Biomedical Optics*, vol. 18, no. 6, p. 066010, 2013.

[101] D. M. Burmeister, A. Ponticorvo, B. Yang, S. C. Becerra, B. Choi, A. J. Durkin, and R. J. Christy, "Utility of spatial frequency domain imaging (sfdi) and laser speckle imaging (lsi) to non-invasively diagnose burn depth in a porcine model," *Burns*, vol. 41, no. 6, pp. 1242–1252, 2015.

[102] A. Yafi, F. K. Muakkassa, T. Pasupneti, J. Fulton, D. J. Cuccia, A. Mazhar, K. N. Blasiole, and E. N. Mostow, "Quantitative skin assessment using spatial frequency domain imaging (sfdi) in patients with or at high risk for pressure ulcers," *Lasers in Surgery and Medicine*, vol. 49, no. 9, pp. 827–834, 2017.

[103] S. Tabassum, Y. Zhao, R. Istfan, J. Wu, D. J. Waxman, and D. Roblyer, "Feasibility of spatial frequency domain imaging (sfdi) for optically characterizing a preclinical oncology model," *Biomedical Optics Express*, vol. 7, no. 10, pp. 4154–4170, 2016.

[104] A. Mazhar, S. Dell, D. J. Cuccia, S. Gioux, A. J. Durkin, J. V. Frangioni, and B. J. Tromberg, "Wavelength optimization for rapid chromophore mapping using

spatial frequency domain imaging," *Journal of Biomedical Optics*, vol. 15, no. 6, p. 061716, 2010.

[105] M. Schmidt, E. Aguénounon, A. Nahas, M. Torregrossa, B. J. Tromberg, W. Uhring, and S. Gioux, "Real-time, wide-field, and quantitative oxygenation imaging using spatiotemporal modulation of light," *Journal of Biomedical Optics*, vol. 24, no. 7, p. 071610, 2019.

[106] J. Angelo, C. R. Vargas, B. T. Lee, I. J. Bigio, and S. Gioux, "Ultrafast optical property map generation using lookup tables," *Journal of Biomedical Optics*, vol. 21, no. 11, p. 110501, 2016.

[107] J. Vervandier and S. Gioux, "Single snapshot imaging of optical properties," *Biomedical Optics Express*, vol. 4, no. 12, pp. 2938–2944, 2013.

[108] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[109] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[110] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[111] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in*

Neural Information Processing Systems, 2014, pp. 2672–2680.

[112] E. Aguénounon, F. Dadouche, W. Uhring, and S. Gioux, "Single snapshot of optical properties image quality improvement using anisotropic two-dimensional windows filtering," *Journal of Biomedical Optics*, vol. 24, no. 7, p. 071611, 2019.

[113] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[114] T. M. Quan, D. G. Hildebrand, and W.-K. Jeong, "Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics," *arXiv Preprint arXiv:1612.05360*, 2016.

[115] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," *arXiv Preprint arXiv:1802.05957*, 2018.

[116] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[117] F. Mahmood and N. J. Durr, "Deep learning and conditional random fields-based depth estimation and topographical reconstruction from conventional endoscopy," *Medical Image Analysis*, vol. 48, pp. 230–243, 2018.

[118] J. Kress, D. J. Rohrbach, K. A. Carter, D. Luo, C. Poon, S. Aygun-Sunar, S. Shao, S. Lele, J. F. Lovell, and U. Sunar, "A dual-channel endoscope for quantitative imaging, monitoring, and triggering of doxorubicin release from liposomes in living mice," *Scientific Reports*, vol. 7, no. 1, pp. 1–13, 2017.

[119] A. K. Dunn, H. Bolay, M. A. Moskowitz, and D. A. Boas, "Dynamic imaging of cerebral blood flow using laser speckle," *Journal of Cerebral Blood Flow & Metabolism*, vol. 21, no. 3, pp. 195–201, 2001.

[120] D. A. Boas and A. K. Dunn, "Laser speckle contrast imaging in biomedical optics," *Journal of Biomedical Optics*, vol. 15, no. 1, p. 011109, 2010.

[121] P. Jain and S. E. Sarma, "Measuring light transport properties using speckle patterns as structured illumination," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.

[122] L. I. Goldfischer, "Autocorrelation function and power spectral density of laser-produced speckle patterns," *Josa*, vol. 55, no. 3, pp. 247–253, 1965.

[123] J. W. Goodman, *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[124] P. Di Ninni, F. Martelli, and G. Zaccanti, "The use of india ink in tissue-simulating phantoms," *Optics Express*, vol. 18, no. 26, pp. 26 854–26 865, 2010.

[125] S. Prahl, "Tabulated molar extinction coefficient for hemoglobin in water. oregon medical laser center," 1998.

[126] S.-H. Tseng, P. Bargo, A. Durkin, and N. Kollias, "Chromophore concentrations, absorption and scattering properties of human skin in-vivo," *Optics Express*, vol. 17, no. 17, pp. 14 599–14 617, 2009.

[127] S. C. Kanick, D. M. McClatchy, V. Krishnaswamy, J. T. Elliott, K. D. Paulsen, and B. W. Pogue, "Sub-diffusive scattering parameter maps recovered using

wide-field high-frequency structured light imaging," *Biomedical Optics Express*, vol. 5, no. 10, p. 3376, Oct. 2014.

[128] M. B. Applegate and D. M. Roblyer, "High-speed spatial frequency domain imaging with temporally modulated light," *Journal of Biomedical Optics*, vol. 22, no. 7, p. 076019, 2017.

[129] T. L. Bobrow, F. Mahmood, M. Inserni, and N. J. Durr, "Deeplsr: a deep learning approach for laser speckle reduction," *Biomedical Optics Express*, vol. 10, no. 6, pp. 2869–2882, 2019.

[130] V. Parot, D. Lim, G. González, G. Traverso, N. S. Nishioka, B. J. Vakoc, and N. J. Durr, "Photometric stereo endoscopy," *Journal of biomedical optics*, vol. 18, no. 7, p. 076017, 2013.

[131] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1. IEEE, 1999, pp. 125–131.

[132] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1. IEEE, 2003, pp. I–I.

[133] J. Angelo, C. R. Vargas, B. T. Lee, I. J. Bigio, and S. Gioux, "Ultrafast optical property map generation using lookup tables," *Journal of Biomedical Optics*, vol. 21, no. 11, p. 110501, 2016.

[134] A. N. Bashkatov, E. A. Genina, V. I. Kochubey, V. Rubtsov, E. A. Kolesnikova, and V. V. Tuchin, "Optical properties of human colon tissues in the spectral

range," *Quantum Electronics*, vol. 44, no. 8, p. 779, 2014.

[135] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, 2005, pp. 807–814.

[136] S. K. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching," in *International Conference on Computer Vision Systems*. Springer, 2009, pp. 134–143.

[137] D. Khan, M. A. Shirazi, and M. Y. Kim, "Single shot laser speckle based 3d acquisition system for medical applications," *Optics and Lasers in Engineering*, vol. 105, pp. 43–53, 2018.

[138] M. Schaffer, M. Grosse, and R. Kowarschik, "High-speed pattern projection for three-dimensional shape measurement using laser speckles," *Applied optics*, vol. 49, no. 18, pp. 3622–3629, 2010.

[139] A. W. Stark, E. Wong, D. Weigel, H. Babovsky, T. Schott, and R. Kowarschik, "Subjective speckle suppression in laser-based stereo photogrammetry," *Optical Engineering*, vol. 55, no. 12, p. 121713, 2016.

[140] U. Upadhyay, V. P. Sudarshan, and S. P. Awate, "Uncertainty-aware gan with adaptive loss for robust mri image enhancement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3255–3264.

[141] H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun, "A survey on deep learning techniques for stereo-based depth estimation," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, 2020.

[142] Z. Li, X. Liu, N. Drenkow, A. Ding, F. X. Creighton, R. H. Taylor, and M. Unberath, "Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 6197–6206.

[143] J. Qian, S. Feng, T. Tao, Y. Hu, Y. Li, Q. Chen, and C. Zuo, "Deep-learning-enabled geometric constraints and phase unwrapping for single-shot absolute 3d shape measurement," *Apl Photonics*, vol. 5, no. 4, p. 046105, 2020.

[144] M. van de Giessen, J. P. Angelo, and S. Gioux, "Real-time, profile-corrected single snapshot imaging of optical properties," *Biomedical Optics Express*, vol. 6, no. 10, pp. 4051–4062, 2015.

[145] H. N. Le, H. Nguyen, Z. Wang, J. Opfermann, S. Leonard, A. Krieger, and J. U. Kang, "Demonstration of a laparoscopic structured-illumination three-dimensional imaging system for guiding reconstructive bowel anastomosis," *Journal of biomedical optics*, vol. 23, no. 5, p. 056009, 2018.

[146] D. Lim, T. N. Ford, K. K. Chu, and J. Metz, "Optically sectioned in vivo imaging with speckle illumination hilo microscopy," *Journal of biomedical optics*, vol. 16, no. 1, p. 016014, 2011.

[147] L.-H. Yeh, L. Tian, and L. Waller, "Structured illumination microscopy with unknown patterns and a statistical prior," *Biomedical optics express*, vol. 8, no. 2, pp. 695–711, 2017.

Curriculum Vitae **Mason Tianyi Chen** **March, 2022**

mason@jhmi.edu

EDUCATION AND DEGREES

2017–Present Graduate student, Department of Biomedical Engineering

Johns Hopkins University

2013–2017 Undergraduate student, Department of Biomedical Engineering

Vanderbilt University

RESEARCH EXPERIENCE

Computational Biophotonics Lab – Dr. Nicholas Durr

Johns Hopkins University, Baltimore, MD *(August 2017 - Present)*

Graduate Research Assistant

- Design an endoscopic imaging system that is co-optimized with artificial intelligence algorithms for improved diagnostic performance.
- Develop physical and machine learning models for automated tissue classification and real-time, quantitative imaging of tissue optical signatures in endoscopy.
- Design a laparoscope and propose a deep learning framework for rapid depth estimation and oxygenation measurements.
- Implement deep generative models for image reconstruction and diabetic retinopathy detection with a lensless, computational funduscope.
- Prototype optical imaging systems and perform simulations to demonstrate the use of polarized and structured light in quantitative characterization of biological tissues.

Focused Ultrasound Lab – Dr. William Grissom

Vanderbilt University, Nashville, TN *(October 2015 - May 2017)*

Research Assistant

- Designed and conducted experiments using background-oriented Schlieren (BOS) imaging for high-intensity focused ultrasound beam mapping and developed a remotely-controlled Android app for BOS imaging.
- Performed focused ultrasound pressure field simulations with cluster computing.
- Investigated the heating and mechanical effect of high-intensity focused ultrasound on drug uptake of tumors using MRI-guided methods.

PROFESSIONAL EXPERIENCE

Amazon.com Services, Inc.

Seattle, WA *(Summer 2021)*

Research Scientist Intern

Spectrum Innovations, LLC

Nashville, TN *(Summer 2015)*

Consultant Intern

TEACHING EXPERIENCE

Johns Hopkins University *(January 2021 - May 2021)*

Teaching Assistant

- EN.580.464 Advanced Data Science for Biomedical Engineering (Instructor: Prof. Brian Caffo)

JOURNAL PUBLICATIONS

Speckle illumination spatial frequency domain imaging for projector-free optical property mapping. **M. T. Chen**, M. Papadakis and N. J. Durr. *OSA Optics Letters* 46, 673-676 (2021).

Rapid tissue oxygenation mapping from snapshot structured-light images with adversarial deep learning. **M. T. Chen** and N. J. Durr. *SPIE Journal of Biomedical Optics* 25(11), 112907 (2020).

GANPOP: Generative Adversarial Network Prediction of Optical Properties from Single Snapshot Wide-field Images. **M. T. Chen**, F. Mahmood, J. A. Sweer and N. J. Durr. *IEEE Transactions on*

Medical Imaging 39(6), 1988-1999 (2019).

Wide-field optical property mapping and structured light imaging of the esophagus with spatial frequency domain imaging. J. A. Sweer, **M. T. Chen**, K. J. Salimian, R. J. Battafarano and N. J. Durr. *Journal of Biophotonics* 12(9), e201900005 (2019).

CONFERENCE PUBLICATIONS

Real-time oxygenation mapping from structured light imaging with deep learning. **M. T. Chen** and N. J. Durr. *Optical Tomography and Spectroscopy of Tissue XIV. SPIE Photonics West* (Oral presentation, 2021).

Towards SFDI endoscopy with structured illumination from randomized speckle patterns. **M. T. Chen**, T. L. Bobrow and N. J. Durr. *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XIX. SPIE Photonics West* (Oral presentation, 2021).

Quantitative Tissue Property Measurements with Structured Illumination and Deep Learning. **M. T. Chen** and N. J. Durr. *IEEE Photonics Conference (IPC)* pp. 1-2 (2020).

Detection and Segmentation of Colorectal Polyps with Fully Convolutional Neural Networks. **M. T. Chen**, G. N. McKay, and N. J. Durr. *IEEE International Symposium on Biomedical Imaging (ISBI)* (2018).

SERVICE AND LEADERSHIP

BME Equality, Diversity, and Inclusivity Committee

Johns Hopkins University, Baltimore, MD *(November 2020 – present)*

Peer Mentor

BME PhD Council

Johns Hopkins University, Baltimore, MD *(August 2018 – August 2019)*

Secretary

Graduate Representative Organization

Johns Hopkins University, Baltimore, MD *(August 2017 – August 2018)*

Department Representative

SPIE Journal of Biomedical Optics

Reviewer

IEEE Transactions on Medical Imaging

Reviewer