Advances in Stimulated Raman Scattering Microscopy via Deep Learning

Bryce Adrian Manifold

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Dan Fu, Chair

Ashleigh Theberge

Dustin Maly

Program Authorized to Offer Degree:

Department of Chemistry
University of Washington

Abstract

Advances in Stimulated Raman Scattering Microscopy via Deep Learning

Bryce Adrian Manifold

Chair of the Supervisory Committee:
Dan Fu
Department of Chemistry

Stimulated Raman scattering (SRS) microscopy is a powerful chemical imaging technique that acquires images based on the vibrational-spectral “fingerprints” of molecules within an imaged field of view often without the need for exogenous fluorophores or labels. SRS microscopy has found an established niche in biophotonics with many examples of translational clinical applications and demonstrations of imaging various biological systems on subcellular to tissue spatial orders. Concurrent to the development of SRS microscopy, computational advancements have seen a democratized adoption of deep learning platforms for a wide variety of computer vision tasks. In this work I document my contributions in integrating SRS microscopy and deep learning towards advancing the capability to study biological systems. Specifically, deep learning will be shown to address technical limitations of SRS microscopy such as imaging noise and ultimate imaging depth in tissue samples. Deep learning will also be shown to improve analysis of SRS images via the development of a novel convolutional neural network architecture designed to handle a variety of chemical imaging techniques and perform a variety of computer vision tasks. Finally, I will show how these advancements and novel architecture can be used to diagnose

thyroid cancer in label-free human tissue samples and to classify and study T cells based on label-free images.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank first my fellow group members for their friendship and collaboration in my graduate studies. Dr. Andrew T. Francis, Dr. Benjamin Figueroa Jr., and Dr. Andrew H. Hill for their assistance, guidance, and mentorship. Elena Thomas, Fiona Xu, and Ruoqian Rachel Hu for their many contributions to the work presented here and their friendship. I also thank Shuaiqian Helen Men and Eleanor Munger for their contributions in maintaining and preparing many biological samples presented here.

I would also like to thank those who provided assistance on these works outside of the Fu group. Drs. Gregory Johnson and Chek Ounkomol for their development of and guidance on the pytorch_fnet codebase used here. Erika French, Kathy Andrich, and the Animal Use Training Services for their many donations of murine tissues used in this work. Byron Tardif, Dr. Amanda Hummon, and Dr. Ariel Rokem for their helpful discussions. Aldous Subang, Piper Driskell, and NWBiospecimen for procurement of human tissues. Eric Huang for his collaboration and guidance in thyroid pathology. Elisa Clark and Professor Hao Yuan Kueh for their collaboration and guidance in the T cell work.

I thank my committee members Professors Ashleigh Theberge, Dustin Maly, Jonathan Liu, and Joshua Vaughan for their guidance and commentary on the completion of this work.

Finally, I thank Professor Dan Fu for his tutelage and training over the past 5 years. I can only hope that under his mentorship I have gained a fraction of his vast knowledge, intellectual incisiveness, and unending perseverance.

DEDICATION

To my darling Allison, you are the light of my life by which all else is visible.

Chapter 1. INTRODUCTION

1.1 MOTIVATION AND OVERVIEW

Stimulated Raman scattering (SRS) microscopy is a powerful optical imaging technique that uses the intrinsic vibrational contrast of molecules to provide chemical maps of biological cells and tissues. Due to its label-free imaging capability and subcellular spatial resolution, SRS imaging has shown great promise in many biological and biomedical applications such as metabolic studies, drug imaging, tissue diagnosis [1–8].

Despite recent advances, SRS imaging still faces technical shortcomings that prevent it from becoming more widely used in biological imaging. Similar to multiphoton fluorescence, SRS imaging uses ultrashort laser pulses to excite weak nonlinear optical transitions [9]. Specifically, two ultrafast laser pulses (pump and Stokes) are used to coherently excite a Raman vibrational transition via a virtual state at the focus of a scanning laser microscope. As the lasers are scanned through the focus of the objective, an image is formed that contains quantitative chemical information at micron scale. However, SRS microscopy sensitivity can be low, especially when trying to image low concentration molecules, deep into tissue, or at high speed. Low signal conditions may also be inevitable when considering the limitation of laser power for *in vivo* imaging to avoid tissue damage [10,11]. *In vivo* experiments with SRS imaging are also often inherently signal-limited where “epi” detection (where weak back-scattered light is collected from the sample) is typically required. To this end we show the capability of deep learning to significantly improve image quality and even extend the ultimate imaging depth limit in scattering tissue samples.

In addition to these technical limitations (imaging depth, laser power, or detection scheme) that affect image quality, SRS imaging may also present limitations in chemical specificity even when used in a "hyperspectral" capacity. That is, where multiple spectral components of a field of view are depicted in their own image. Though the vibrational information afforded by SRS microscopy can be specific to a given molecule, there are often many overlapping contributions to vibrational signals, especially in biological samples, that confound image interpretation. This may belie claims of the value of SRS as a label-free imaging technique, as any value of avoiding labels may be offset by not knowing exactly what molecular vibration is being probed. Linear decomposition, phasor analysis, and some machine learning methods have been developed for analysis of hyperspectral imaging datasets[12–16], however these methods tend to focus on interpretation of only the spectral information while the spatial information may be underutilized. Recently, deep learning via Convolutional Neural Networks (CNNs) has shown significant promise in both enhancing image quality [17–19] and improving hyperspectral image analysis [20–24]. Indeed, with the continuous improvement and increasing availability of advanced computing hardware and open-source methods, deep learning is finding broader use in a wide variety of imaging, sensing, and biophotonics research [25,26].

In this work, my contributions of using deep learning to improve SRS image quality and hyperspectral image analysis will be presented. Specifically, deep learning denoising is shown to significantly improve SRS image quality in differing sample types and to be applicable to SRS images outside of the explicitly trained data space [27]. I will also present corollary work better defining the limits of SRS imaging depth in tissue and how such depths can be extended via deep learning [28]. I will then present my work in creating and utilizing a novel deep learning architecture (the U-within-U-Net, or UwU-Net) for hyperspectral image analysis [29].

Specifically, I will show the ability of the UwU-Net to perform simultaneous segmentation and classification of hyperspectral remote sensing imagery and label-free prediction of fluorescence images from hypserpectral SRS images. I then show how these advancements in deep learning for SRS microscopy can be translated both to clinical relevance by predicting the diagnosis of thyroid cancer samples, and to study biological systems such as immunological cells in a label-free capacity.

1.2 Stimulated Raman Scattering Microscopy

SRS is a third-order nonlinear interaction that coherently probes the same vibrational chemical information recovered from a linear (or spontaneous) Raman experiment[9]. This information arises from the vibrational excitation of electrons within bonds to higher energy vibrational states via a virtual excited state (Figure 1.1) induced by an photon of some energy that corresponds to neither a real vibrational nor electronic eigenstate. This stands in contrast to linear infrared absorption where photons of energies equal to some vibrational transition (these wavelengths are in the infrared section of the electromagnetic spectrum) are directly absorbed to acquire vibrational information (Figure 1.1). Given infrared photons' strong water absorption cross section, linear infrared spectroscopy is usually not ideal for a microscopy of biological systems. Instead Raman scattering has found a particular niche here with fast and readily quantitative analysis of vibrational chemical information.

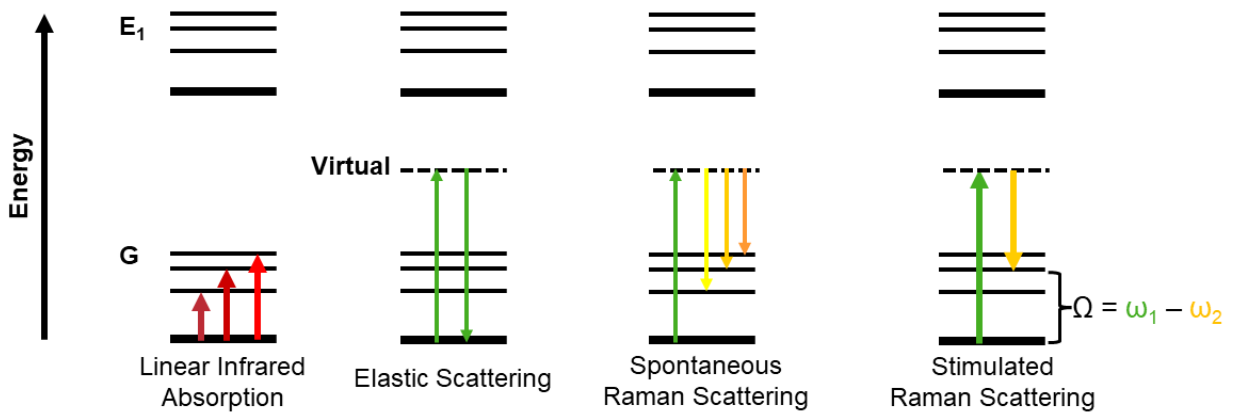

Figure 1.1. Energy Level Diagrams of Vibrational Spectroscopy Interactions

The energy level diagrams represent some theoretical molecule with electrons trapped in bonds. The solid black lines are shown grouped into 2 electronic states denoted by "G" and "E_1" representing the ground and 1st electronically excited state, respectively. Within each electronic state are shown vibrational energy levels. Photons are represented as arrows whose lengths correspond to their respective energies. Different energies of photons are color coded for clarity. The vibrational energy levels can be excited directly via linear infrared absorption (left most) or by a Raman process (either spontaneous or stimulated). Elastic scattering is also shown.

Stimulated Raman signal, specifically, is generated by the coincident interaction of two laser pulses (pump and Stokes beams) at the sample. There, a pump-probe process occurs by which the probed Raman band of frequency, Ω, corresponds to the difference in frequency of the pump and Stokes beams, i.e. $\Omega = \omega_{pump} - \omega_{Stokes}$ (Figure 1.1). The three-photon interaction and resulting signal photon can be thought of as transmuting a pump photon into a Stokes photon via the coupled Raman band. This intensity loss in the pump pulse is stimulated Raman loss (SRL) and the subsequent gain in intensity of the Stokes pulse is stimulated Raman gain (SRG). In this work, only the SRL process is discussed as it is the detection scheme used for all experiments. The SRL process only attenuates the pump beam by a miniscule fraction often far below the typical pulse-to-pulse intensity differences. In order to quantify the transfer of energy (i.e. number of nonlinear

events) between pulse trains, lock-in detection is employed. By modulating the Stokes pulse train (typically 10-40 MHz) and detecting the modulation transfer to the pump pulse train, SRS signal can be detected down to millimolar sensitivity. If the pump and Stokes beams are raster scanned through a microscope with known speed and subsequently detected at a photodiode triggered to the scanning rate, an image can be constructed with pixel values corresponding to SRS signal.

Stimulated Raman scattering (SRS) microscopy was first demonstrated in 2008 by Freudiger et al. as an alternative coherent Raman microscopy technique to the then-more-popular Coherent Anti-Stokes Raman Scattering (CARS) microscopy [30]. Both SRS and CARS microscopy use synchronized pulsed lasers to coherently excite the inherent chemical vibrations of molecules. This contrasts with so-called "spontaneous Raman scattering", where a single (typically continuous wave) laser source focused on a sample will cause spontaneous emission of photons of different wavelengths (i.e. energies) corresponding to vibrational energy differences with respect to the original laser source. While spontaneous Raman scattering affords broad and readily interpretable spectral information about a sample, the probability of spontaneously emitted Raman photons is typically quite low with differential scattering cross-sections in the ~10^{-30} cm^2 regime [31–33]. Thus, if microscopy is desired via spontaneous Raman, acquisition times are often egregiously slow given the typical excitation wavelengths and maximum laser powers appropriate for biological samples. Coherent Raman scattering microscopies offer significantly faster and more sensitive Raman imaging via coherent nonlinear optical interactions typically at the expense of broad spectral coverage and some spectral specificity [34,35].

CARS microscopy has seen broad use in biophotonics research since its first demonstration but has largely fallen wayside to SRS microscopy [36–42]. This is due largely to the shortcomings of CARS microscopy with respect to SRS microscopy. Specifically, CARS microscopy, though

technically easier to implement and detect signal, suffers from a significant “non-resonant” background effect and spectral distortion (with respect to the spontaneous Raman spectrum). This non-resonant background and resultant spectral distortions (wavenumber shifts, peak broadening, peak intensity changes, etc.) make quantitative chemical analysis of CARS microscopy images difficult [43,44].

SRS microcopy, in contrast, does not suffer from the non-resonant background issue. SRS signal arises from energy transfer between the two lasers sources and the molecular vibration, with spectral features nearly identical to spontaneous Raman except for slight peak broadening related to the pulse duration and bandwidth of the lasers [9]. Additionally, SRS microscopy offers the distinct advantage of signal strength that is linearly dependent on sample concentration. This affords SRS microscopy with fast, sensitive, label-free, and quantitatively intuitive detection of chemically specific imaging contrast of samples. Moreover, SRS microscopy is well-suited for biological samples because Raman scattering of water vibrations is relatively weak and the typically used near infrared picosecond laser sources cause minimal photodamage due to heating and multiphoton ionization. Indeed, SRS microscopy has found its largest use in biophotonics as a label-free yet chemically specific alternative to more ubiquitously adopted fluorescence microscopy techniques. While fluorescence microscopy is a valuable tool with relatively low barrier to utilize, the use of exogenous fluorophores has some commonly cited shortcomings [35,45–49], including label specificity, cytotoxicity, photobleaching, and color cross-talk. SRS microscopy has found utility in a wide variety of biological applications such as imaging pharmacokinetics [50–54], cellular and physiological metabolism [55–59], antibiotic susceptibility and resistance [60,61], and histopathological diagnosis [62–65]. While such experiments could

also be performed using fluorescent labels, the value of a label-free method that also provides quantitative insight is highly desirable for modern biological studies.

However, this quantitative insight is limited in other ways. For example, scattering of light in turbid media (as biological samples often are) often makes calibration and validation of quantitative SRS measurements difficult [66]. Further, because the SRS process and detection scheme relies on focusing ultrafast laser pulses through a high numerical aperture objective, other nonlinear processes often confound or distort SRS signal. For example, given the presence of appropriate molecules, transient absorption (absorption of multiple photons to excite an electronic transition of molecules) processes may also occur at the focus of the objective giving rise to spurious signal features unrelated to the desired vibrational information. Cross-phase modulation (XPM) is often another parasitic process that contributes broad chemically-nonspecific background signal in SRS microscopy. XPM arises from a nonlinear modulation of refractive index at the focus of the objective causing down field focus changes at the same frequency as the SRS signal being sought by the lock-in amplifier. Similarly, thermal lensing, where a temperature gradient creates a similar effect, can also contribute background signal in a resultant SRS image. Indeed, the background limitation of SRS has warranted many technical improvements to improve the sensitivity of SRS [8,67].

These background limitations are often related to the pulse duration used in SRS microscopy. Modern SRS microscopes typically are operated in two main types in this regard based on the laser source chosen. The first type utilizes an ultrafast source with output pulses in the 5-12 picosecond regime. This allows for higher average power at focus (as longer duration pulses have lower peak energies meaning less sample damage) and higher spectral resolution (i.e. the ability to resolve nearby peaks in the vibrational spectrum). The trade off is slightly lower

signal at commensurate average powers, and slower acquisition of hyperspectral images (where images are taken at different Raman transitions) due to having to tune the laser wavelength output at the cavity. The second type of system, and that which is used throughout all of the SRS experiments presented here utilizes a broadband femtosecond source. Where the pulses leaving the laser cavity near the transform limit (~100 fs) with ~10 nm bandwidth at the center wavelength. This offers slightly higher signal and quick acquisition of hyperspectral data (when used in the spectral focusing mode). The trade offs here are potentially more damage being done to samples, typically lower spectral resolution and the above mentioned background issues.

Ultimately it is important to consider the comparison of SRS microscopy to other similar microscopy methods, especially fluorescence microscopy which stands as a gold standard method for subcellular microscopy. Stated plainly, SRS microscopy achieves 3 main advantages over fluorescence microscopy:

1. Label-free visual contrast based on endogenous molecules at near equivalent spatial resolution
2. No photobleaching, as the molecules providing the visual contrast are not breaking down during continued imaging.
3. Inherently quantitative information given appropriate calibration and validation

But SRS microscopy also suffers from significant trade-offs in comparison to fluorescence microscopy:

1. Lower sensitivity, usually μM sensitivity at best versus all the way down to single molecule detection
2. Scattering limits imaging depth to 100's of μm versus multiple mm.

3. Specificity issue when multiple endogenous molecules contribute SRS signal at a given Raman transition, as is common in the carbon-hydrogen region. This additionally complicates interpretation of hyperspectral image datasets

Throughout this work I present how the challenges posed by these limitations in SRS microscopy can be addressed through deep learning, and how deep learning can augment the capabilities and applications of SRS microscopy.

Chapter 2. DEEP LEARNING DENOISING OF STIMULATED RAMAN SCATTERING MICROSCOPY IMAGES

The work presented in this chapter has been published in the following article[27]:

Manifold, B.; Thomas, E.; Francis, A. T.; Hill, A. H.; Fu, D. "Denoising of Stimulated Raman Scattering Microscopy Images via Deep Learning" Biomed. Opt. Express, BOE 2019, 10 (8), 3860–3874

2.1 INTRODUCTION

As described in the previous chapter, SRS images, though powerful towards imaging biological samples in a label-free modality, often suffers with respect to sensitivity with respect to the more gold-standard fluorescence imaging. Indeed, typical SRS images often are noisy and of low quality when imaging deep and/or at high speed. For example, SRS signal is often unacceptably noisy when imaging tissue at depths below 90 μm even for high abundance biomolecules such as proteins and lipids [68,69]. Additionally, low signal conditions may be inevitable when compensating for potential damage to biological samples when imaging at short pulse durations or high powers. *In vivo* experiments with SRS imaging may also be signal-limited where "epi" detection is typically required in comparison to experiments where light can be transmitted through a sample and collected on the other side. Epi imaging acquires back-scattered light from the sample which yields significantly weaker signal in comparison to transmission mode detection unless samples are sufficiently thick (as will be discussed in greater detail in Chapter 3). These challenges (depth, laser power, and detection scheme) are common in biological imaging and often result in the acquisition of low SNR images. While in some applications this can be mitigated by increasing imaging times, *in vivo* applications require rapid image collection to both

observe physiologically relevant processes and avoid imaging artifacts due to sample motion. As a result, many *in vivo* studies turn to denoising algorithms to increase their image SNR [70–72].

While standard denoising algorithms can be used to improve image quality, they typically require either *a priori* knowledge about the interfering noise or multiple images of the same features to enable averaging and often introduce undesirable consequences such as a decrease in the effective spatial resolution of the image [73–75]. Recently, deep learning via CNNs has shown significant promise as a denoising tool [17–19]. These CNN-based algorithms have been used to denoise images with inherent compression corruption or induced Gaussian noise, often even performing well in blind denoising tests. However, the fully-connected architecture of the most common CNNs for denoising involve significant training times and require large training samples to be effective. Moreover, these deep learning denoising algorithms are based on RGB images with relatively narrowband noise (noise centered around a small frequency range) [76]. To our knowledge, this is the first report of using a CNN to denoise nonlinear optical images, particularly SRS images.

Here, we report the use of a U-Net architecture CNN to denoise SRS images in low signal situations. Previous work with this U-Net architecture has been able to create algorithms that predict label-free fluorescence images from brightfield microscopy images with high fidelity while requiring relatively few training images [77]. The use of a CNN presents an elegant way of tailoring a specialized denoising algorithm to significantly improve the quality of SRS images in situations where low signal is unavoidable. In this work, we train a deep learning algorithm with corresponding SRS images taken at low and high laser power (i.e. images with low and high signal to noise ratios) then use the trained algorithm to denoise new images with similarly low SNR. Our method significantly outperforms other denoising methods with respect to many common noise

and image fidelity metrics. Moreover, we find that the trained algorithm is applicable to images acquired at differing fields of view, imaging powers, imaging depths, and even experimental geometries (epi versus transmission) than the images used to train the algorithm. Lastly, we note that while the denoising algorithm is demonstrated for SRS imaging, it should be equally applicable to other nonlinear optical imaging techniques, providing a generalizable method to improve tissue imaging quality. Our findings demonstrate the power of CNN-based deep learning as a denoising technique and provide an avenue to significantly improve the quality of biological images acquired in a wide variety of low SNR conditions.

2.2 METHODS

2.2.1 *Sample Preparation*

HeLa cells were cultured in Dulbecco's modified eagle medium with 10% fetal bovine serum at 37 °C with 5% CO_2 atmosphere. Cells were seeded on coverslips 24 hours prior to being fixed using 1% paraformaldehyde.

Murine brain tissue was harvested from recently sacrificed animals provided by UW Animal Use Training Services (AUTS) according to IACUC protocol 3388–03. After excision, thin (~200 μm) sections of tissue were collected and mounted on glass microscope slides.

2.2.2 *SRS imaging*

SRS images were acquired using a homebuilt SRS microscope as described previously [78–80] and as shown in Figure 2.1. The laser used is a femtosecond dual-output Spectra-Physics Insight DeepSee+ which emits a tunable beam (680-1300 nm) and fixed beam 1040 nm pulse trains at a synchronized 80MHz repetition rate. For HeLa cell imaging, a spectral focusing approach, described elsewhere [81], is adopted. The 800 nm pump pulse is chirped using high density glass,

while the Stokes pulse (centered at 1040 nm) is stretched using a grating-based pulse stretcher [82]. HeLa cell images were collected at 2913 cm^{-1} using a spectral resolution of 15 cm^{-1}. In the two-color mouse brain images two 1040 nm pulse trains, modulated 90° out of phase with each other, are used to enable simultaneous two-color acquisition as shown previously [83]. Murine brain images were collected at 2913 cm^{-1} and 2994 cm^{-1} with a spectral resolution of 45 cm^{-1}. The microscope used is a Nikon Eclipse FN1 equipped with a 40x 1.15 NA water immersion objective for HeLa images and a 25x 1.05 NA water immersion objective for the tissue images. In the HeLa cell imaging, the pump beam (800 nm) power was held constant at 20 mW at focus, and images were taken using 1 mW and 20 mW of Stokes power for each field of view. In the mouse brain imaging, the pump beam was held at 20 mW and images were taken at either 1mW or 15mW each for both Stokes beams for all fields of view. All images collected were 512 × 512 pixels.

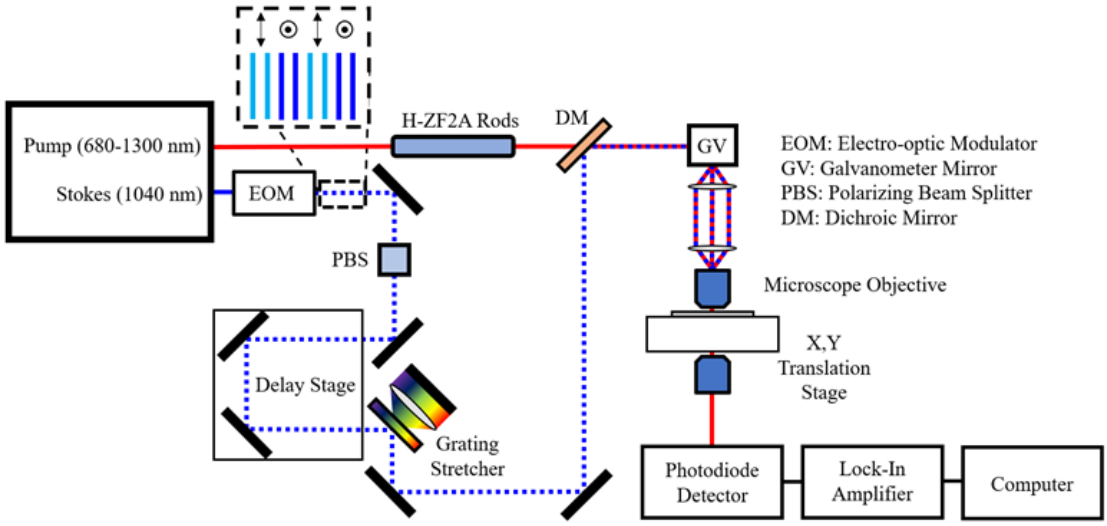

Figure 2.1. SRS microscope setup with spectral focusing

Epi detection used the same setup with the addition of a polarizing beam splitter prior to the objective to collect back scattered light and send it to the photodiode detector.

2.2.3 *Deep Learning Training and Denoising*

In this work, a U-Net CNN was utilized for creating deep learning denoising models[84]. The U-Net CNN creates a series of filters that images get passed through as they are successively broken down into lower resolution components. The filters used to break down the initial image are then subjected to a long training process minimize the mean square error (MSE) between the high-power, low noise truth images and the prediction based on the low-power, high noise images. The expectation is that some filters learn to address "macro" effects (optical aberration, nonuniform illumination, object shape and size, etc.) while others address "micro" effects (Poisson noise, pixel-to-pixel variations, fine structural features, etc.). Those filters accounting for macro effects likely reside higher in the architecture, where the image resolution is still comparatively high and the large portions of the image are considered collectively. Macro effects are sample specific, and much of the structural information utilized in the U-Net CNN training process is likely dealt with in these high level filters. Filters accounting for micro effects likely lie lower in the architecture, where image resolution has been significantly reduced and smaller portions of the initial image are considered. Because of the structural specificity learned by the CNN model, best results are only attainable on a model which has been trained on the specific system it is being asked to denoise. For this reason, different models were trained for each of the systems (HeLa cells and murine brain) studied in this work.

The U-Net CNN used in this work was created by Ounkomol et al[77] with small optimizations made for our specific applications. Corresponding low and high SNR images (i.e. low/high power or epi/transmission modality) were used without any pre-processing to train a denoising algorithm over 50,000 epochs. All deep learning algorithms were supplied 40 fields of view for training with a randomized 10/30 test/train split. The CNN used in this work utilizes a

four-layer network. Each layer consists of two 3x3 kernel convolutions followed by batch normalization and a ReLU activation function, then 2-pixel convolutions followed by batch normalization. Our CNN employs a learning rate of 0.001 with an Adam optimizer, momentum values of 0.5 and 0.999, and a batch size of 20 images.

In the case of the two-color mouse brain images, lipid and protein images were fed simultaneously with additional fields of view withheld for further validation. All images shown here and those used for the relevant peak signal to noise ratio (PSNR), root mean squared error (RMSE), and correlation coefficient (CC) comparisons were not part of the training of the deep learning algorithm. All training sessions and predictions were performed on the University of Washington Hyak Mox supercomputer equipped with an Nvidia P100 GPU. Training sessions lasted ~7 hours depending on training batch and buffer sizes. Utilization of the trained algorithm to denoise batches of images took on average 10 seconds depending on the size of the batch.

2.3 Results and Discussion

2.3.1 *Denoising Low Power SRS Images of Fixed HeLa Cells*

We first denoised images of HeLa cells acquired at 2920 cm^{-1} with low optical power. Because the SNR is linearly proportional to the Stokes power in SRS imaging [85], images acquired at 1 mW (Fig. 2.2A) display 20-fold lower SNR than those acquired at 20 mW (Fig. 2.2D). Low power images appear grainy with low cellular contrast relative to the background and lacking the expected distinct features such as lipid droplets and nuclei as seen in the corresponding high power SRS image. Pairs of corresponding low and high SNR images similar to those shown Fig. 2.2A and 2.2D provided the signal and truth images for the deep learning algorithm training. The resulting algorithm takes an input low SNR image (Fig. 2.2A) and returns a predicted high SNR image such as the image shown in Fig 2C. For comparison, the image shown in Fig. 2.2A is also denoised with

variance stabilization transform (VST) Poisson denoising [86], the result of which is shown in Fig. 2.2B. To better visualize the ability of the denoising techniques to recover cellular features and background contrast, pixel value plots along the same line region in all images are shown in Fig. 2.2E. An image of the same field of view denoised using PURE-LET is provided in the Appendix A (Fig. A1). In the low SNR plot (orange), the variation in pixel value along the image dominates the cellular features more clearly seen in the high SNR line plot (red). While these spatial features are partially recovered and the variation from the noise is suppressed in the VST denoised plot (magenta), the sharp features (cell edges, lipid droplets, nucleoli, etc.) are not well recovered and significantly blurred. The deep learning algorithm, however, demonstrates significant denoising of the low SNR image with near perfect separation of the cells from the background and significant recovery of cellular features such as lipid droplets and nuclei.

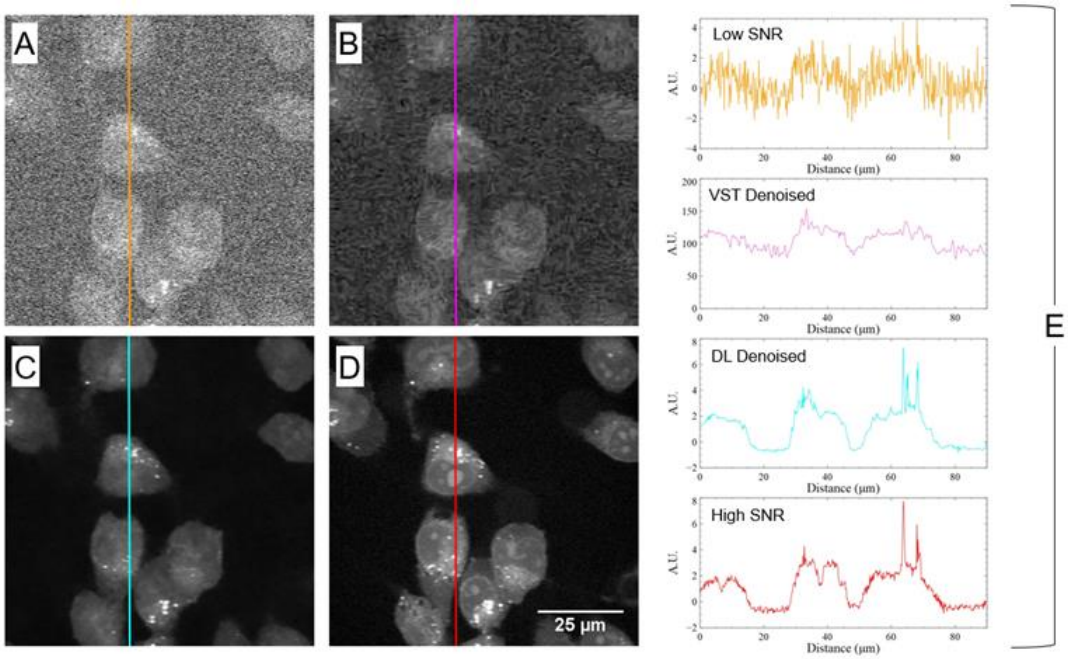

Figure 2.2. Denoising of HeLa Cell SRS Images

(A-D) SRS images at 2920 cm-1 of fixed HeLa cells where (A) was imaged using 1mW Stokes and 20 mW pump (low power) then subsequently denoised using VST (B) and the trained deep

learning algorithm (C); (D) shows the same field of view as (A-C) imaged at 20 mW Stokes and pump (high power). Corresponding pixel value plots along the shown line are demonstrated in (E).

To further quantify the denoising capability of the deep learning algorithm in comparison to other denoising methods, PSNR, RMSE, and CC values were calculated with respect to the high power images (equations and explanations shown in Appendix A). PSNR is a metric that expresses a logarithmic measure of image quality with respect to a truth image. High PSNR indicates higher image fidelity. RMSE expresses the accuracy of the denoising method with respect to a truth image. A low RMSE indicates an accurate denoising method. CC is the Pearson correlation coefficient that expresses colocalization of features in the test and truth images as a number between -1 and 1. A CC of -1, 0, or 1 would indicate perfect anti-correlation, no correlation, or perfect correlation respectively. These values were calculated in ImageJ using previously written plugins [87] with the withheld test images to avoid concerns of overfitting. As shown in Table 2.1, the deep learning denoising significantly outperforms other denoising algorithms optimized for removing Poisson-shaped noise (VST and PURE-LET [86,88]). VST denoising slightly decreases PSNR and increases RMSE in comparison to the original input image likely due to the extremely low starting PSNR and significant blurring of spatial features during the denoising process.

Table 2.1. Denoising Quality Metrics of HeLa Cell SRS Images.

Images/Denoising Method	PSNR (dB)	NRMSE	CC
Low Power	9.99	0.279	0.34
VST Denoised	9.78	0.289	0.59
PURE-LET Denoised	13.53	0.186	0.70
Deep Learning Denoised	**28.87**	**0.049**	**0.94**

*Calculated with respect to high power images. Numbers based on 10 witheld test images

Overall, our data suggests that given the correct training (even on a relatively small training set of 30 images), the U-Net CNN deep learning algorithm demonstrates strong capability for denoising images taken using low powers.

2.3.2 *Denoising Low Power Two-color SRS images of ex vivo Mouse Brain*

To further assess the utility of deep learning denoising in SRS imaging of tissue, we utilized our algorithm to denoise murine coronal brain tissue section images. While images of fixed HeLa cells have well-defined internal features and a truly signal-free background (parameters which U-Net architectures excel in learning [89]), murine brain tissue exhibits significantly more heterogeneity and no true background. In this case our algorithm was trained with corresponding low and high power two-color SRS images acquired in transmission mode.

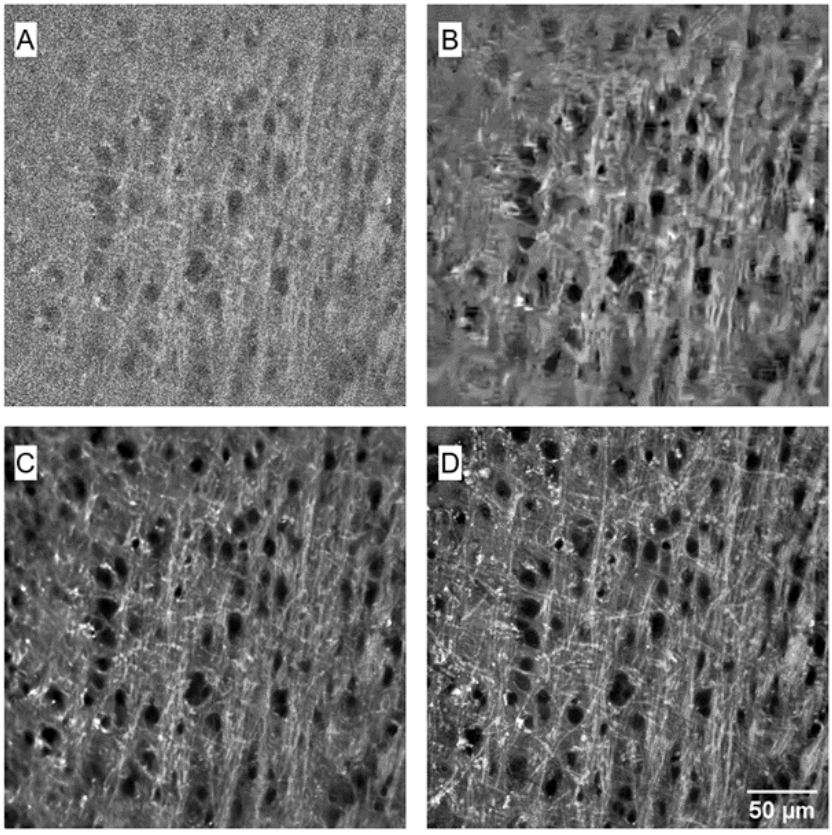

Figure 2.3. Denoising SRS images of mouse brain

Coronal mouse brain SRS images acquired at 2990 cm^{-1}. (A) Low power image acquired at 1 mW Stokes and 20 mW pump. (B) The low power image denoised with VST. (C) The low power image denoised with the deep learning algorithm. (D) The high power image acquired at 20 mW Stokes and 20 mW pump.

The relative denoising capability of the algorithm is first shown in Fig. 2.3 with only lipid channel (2990 cm^{-1}) of the two-color images. The lipid images shown in grey scale demonstrate the high fidelity of denoising by the deep learning-trained algorithm. Figs 3A and 3D show the low and high power images for the field of view respectively. Figs 3B and 3C show the VST and deep learning denoised versions of the low power image respectively. It is clear given the low SNR of the initial image, that VST again significantly blurs spatial features in comparison to the high power image. The deep learning denoised image, however, demonstrates significant denoising without the loss significant loss in spatial resolution, especially in the center of the image. Spatial field of view heterogeneity is common in SRS imaging due to chromatic aberration near the edges of the image. VST denoising fails to recover features around the edges while deep learning does reasonably well.

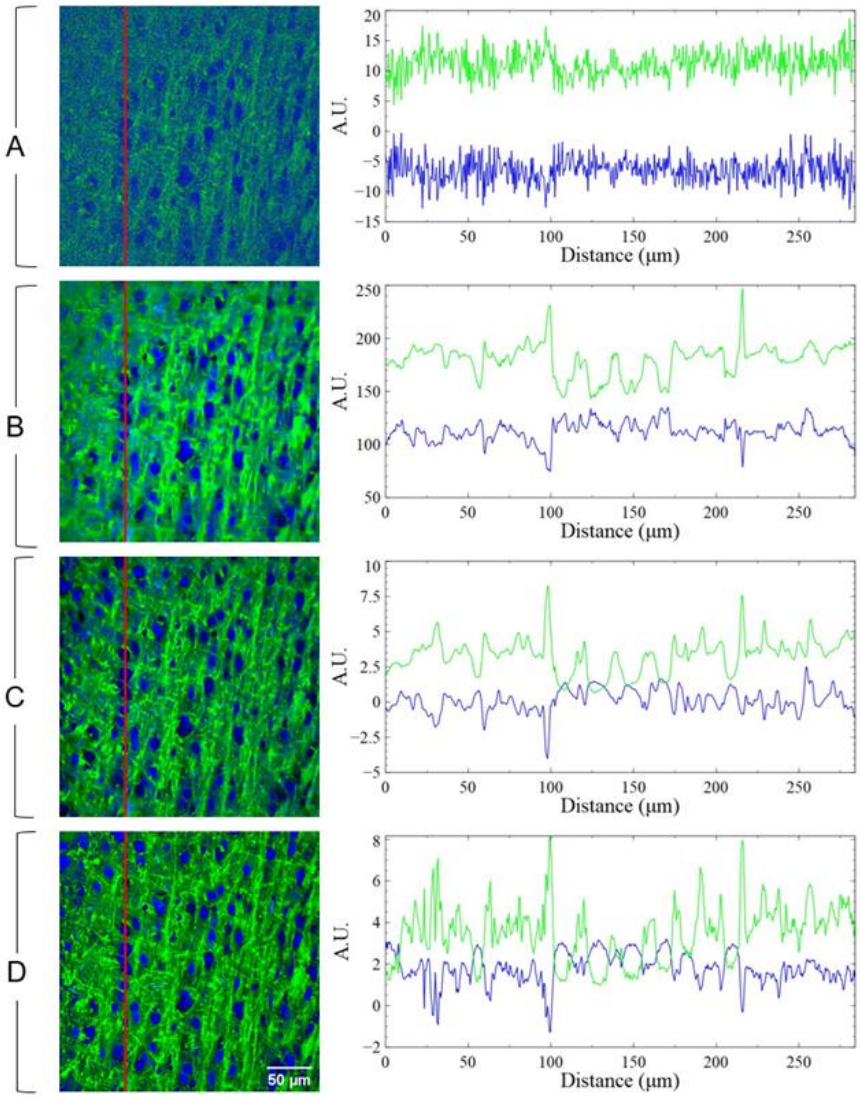

Figure 2.4. Denoising two-color SRS images of mouse brain

Two-color (lipids and proteins) composite SRS images of *ex vivo* mouse brain and corresponding pixel value line plots. (A) Low power image acquired at 1 mW Stokes 20 mW pump. (B) The low power image denoised using VST. (C) The low power image denoised with the deep learning algorithm. (D) The high power image acquired at 20 mW Stokes and pump. Pixel value line plots along the red line are shown for each composite image.

The two-color SRS images (where lipids are colored green and proteins are colored blue) at an imaging depth of ~ 10 μm are shown in Fig. 2.4. In two-color SRS images, significant contrast between the lipid and protein transitions can be used to generate diagnostic maps for pathology

applications [5,90,91]. In the low power images (Fig. 2.4A) however, such contrast is absent, making low SNR images inadequate for pathology. While both VST denoising (Fig. 2.4B) and CNN denoising (Fig. 2.4C) recover the nuclei contrast, the latter performed significantly better in terms of imaging fidelity and spatial resolution. The pixel value plots along the shown lines for both the lipids and protein channels are shown next to their respective images. Here it is evident that, despite the noise in the initial low power image and the heterogeneity in tissue features, deep learning significantly recovers many of the sharp features (mostly axons in green and nuclei in blue) visible in the high power image. While VST does remove a significant portion of the noise, features are blurred, especially around the edges of the image (as seen in Fig. 2.2 and Fig. 2.3). Further, the relatively narrow axons evident in the line plots are recovered with high fidelity indicating deep learning denoising does not sacrifice spatial resolution as other denoising techniques do.

Analysis of PSNR, RMSE, and CC values in these two-color images (shown in Table 2.2) reveals a similar trend to that seen in the HeLa cells. That is, deep learning significantly outperforms other Poisson denoising methods across all metrics for these low power images. While the CC value doesn't approach unity as in the HeLa images, this is likely due to the more heterogeneous background in the tissue compared with the fixed HeLa cell images.

Table 2.2. Denoising Quality Metrics of Mouse Brain SRS Images.

Images/Denoising Method	PSNR (dB)	NRMSE	CC
Low Power	18.04	0.133	0.59
VST Denoised	17.18	0.141	0.73
PURE-LET Denoised	19.75	0.112	0.74
Deep Learning Denoised	**25.24**	**0.066**	**0.82**

*Calculated with respect to high power images. Numbers based on 10 witheld test images including lipid and protein images

This data demonstrates that U-Net based deep learning can create a powerful denoising algorithm with relatively small training sets not only for low power SRS imaging of fixed cells, but also of heterogenous tissue samples. This suggests the possibility of using deep learning to further enhance the current capabilities of SRS imaging.

2.3.3 *Denoising Deep SRS Images of ex vivo Mouse Brain*

One major limitation of deep learning based denoising approaches is blind denoising of images markedly different from those in the training set. This would pose a potential limitation to denoising SRS images deep in tissue as adequate supervised-learning training sets cannot be created due to the inherently low SNR. However, because SRS images at any depth share the same noise features, we hypothesize that we can apply the algorithm trained at shallower imaging depths to images deeper into tissue. We tested this approach by using the algorithm trained in the previous section to SRS images of the same tissue at depths up to 175 μm, a depth that has not been reached before in previous reports of native SRS imaging. In this validation, two-color images were taken at high power (15/20 mW Stokes/pump for both channels) at depths of up to 175 μm into the mouse brain tissue. They were then denoised using the previously shown VST denoising and CANDLE denoising [92]. Images at a depth of 175 μm and denoising of the acquired SRS image are shown in Fig. 2.5. Here, there is no truth high SNR image with which to compare. The SRS image taken at high power is shown in Fig. 2.5A and denoised via VST, CANDLE, and the trained deep learning algorithm are shown in Figs. 5B-5D respectively. The corresponding pixel value line plots are shown for each image. Fig 5A demonstrates the loss in SNR as images are acquired deeper into tissue, even when higher powers are used. From the images and line plots it is clear VST and CANDLE do not fully remove the noise inherent to SRS images at this depth. The deep learning denoising significantly removes the noise, resulting images and pixel plots similar to the denoised

low power images at shallower depths (Figs. 3 and 4). Specifically, the deep learning denoising recovers the expected axons in the lipid channel and better resolves nuclei from the background in the protein channel without significant blurring of any features.

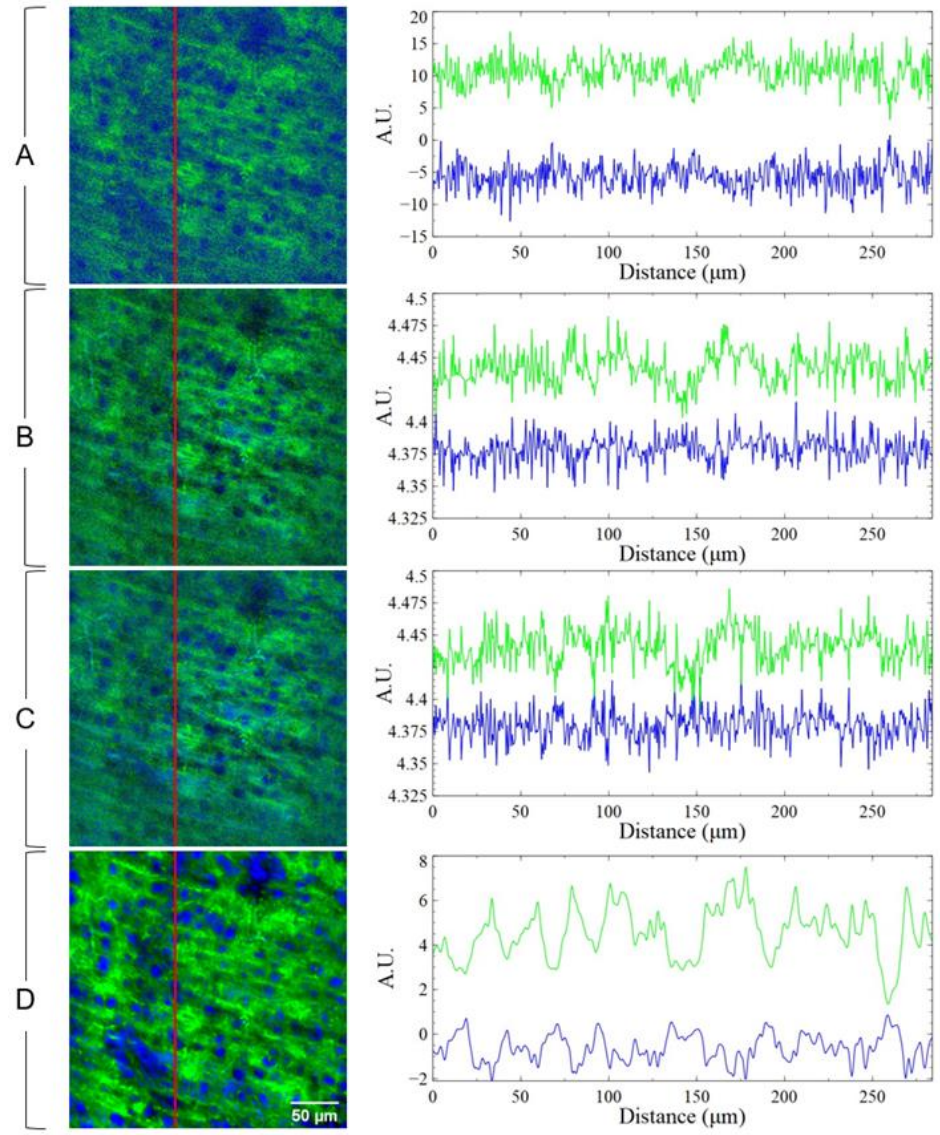

Figure 2.5. Denoising SRS images from 175 μm into brain tissue

(A-D) Two-color (lipids-green, proteins-blue) SRS images of coronal mouse brain slice at a depth of 175μm. (A) shows the field of view taken at 15 mW Stokes 20 mW pump (high power) and (B-D) show denoising of (A) via VST (B), CANDLE (C), and the trained deep learning algorithm (D). Corresponding pixel plots along the lines shown next to their respective images.

The recovery of signal in both the lipid and protein channels demonstrates that the training of the deep learning algorithm at low and high powers in shallower tissue imaging can improve images acquired at much deeper depth. While the laser powers used here are the same as in the high power images shown previously, the images collected have inherently low SNR due to power lost to tissue absorption and scattering. In SRS imaging where shot noise is expected to be the limiting factor in low signal regimes, it would follow that a deep learning algorithm trained to recover signal among this shot-noise would effectively improve imaging depths. The validation shown in Fig. 2.5 suggests that the algorithm is robust in denoising these two-color SRS images, despite not having explicitly trained on images this deep into the brain. Furthermore, it is unnecessary to acquire multiple images of a given field of view to create an average denoised image. CANDLE, for example, requires multiple images over which to learn the average noise distribution. In this case one need only provide a reasonable training set from which the algorithm may learn, then any single image may be denoised using the trained algorithm.

To further examine the generalizability of the algorithm trained in Fig. 2.4, the algorithm was also validated with a low power image taken with a 142 x 142 μm field of view, compared to the 285 x 285 μm fields of view used in the training set. As shown in Fig 2.6A-2.6C, the algorithm remains effective in denoising the low power image even at a zoom different from the images utilized for the training set. The axons and nuclei are reliably recovered without any significant augmentation of shape or size. This further indicates that while the model likely uses structural information in its predictions it does not strictly impose the object size distributions native to its training set. This suggests that a well-trained algorithm would be widely applicable for low SNR situations of a given sample.

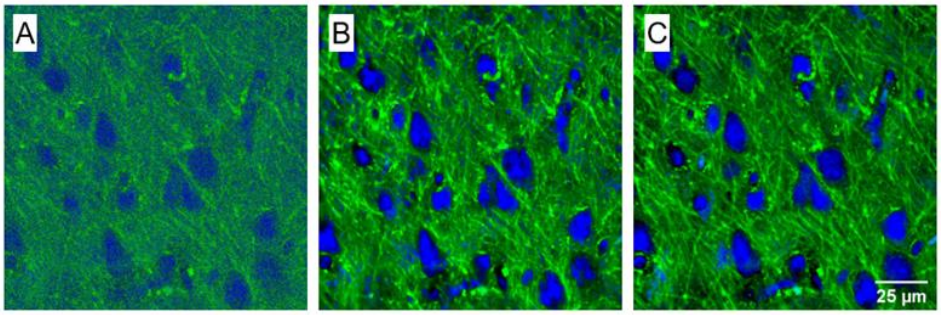

Figure 2.6. Denoising SRS images at 2x zoom. Two-color (lipids-green, proteins-blue) SRS images of coronal mouse-brain slice. Images (A-C) show an area imaged at 2x zoom where (A) was imaged at low power, (B) was the low power image denoised using the same deep learning algorithm from Figs. 3 & 4, and (C) was imaged at high power.

2.3.4 *Denoising epi Two-color SRS Images of Mouse Brain using a Model Trained on Transmission Images*

Finally, we examine the ability of deep learning to create an algorithm that can improve the quality of epi-SRS images to the level of transmission SRS images. This capability is particularly important for *in vivo* imaging, where epi-imaging is required due to the opacity and thickness of many samples. Less light is recovered and directed towards the detector in epi-SRS imaging with respect to the transmissive experiment. This results in a significant loss in SNR. We hypothesize that we can apply training algorithm obtained from transmission images to improve epi SRS imaging and reduce the image quality discrepancy between the two experiments.

Fig. 2.7 depicts the application of our denoising algorithm, trained with transmissive SRS images to the denoising of images collected using an epi-SRS geometry. The images shown in Figs. 2.7A and 2.7C show high power two-color images of murine brain at a depth of 15 μm in epi and transmission mode respectively. Fig. 2.7B shows the epi image denoised by the deep learning algorithm.

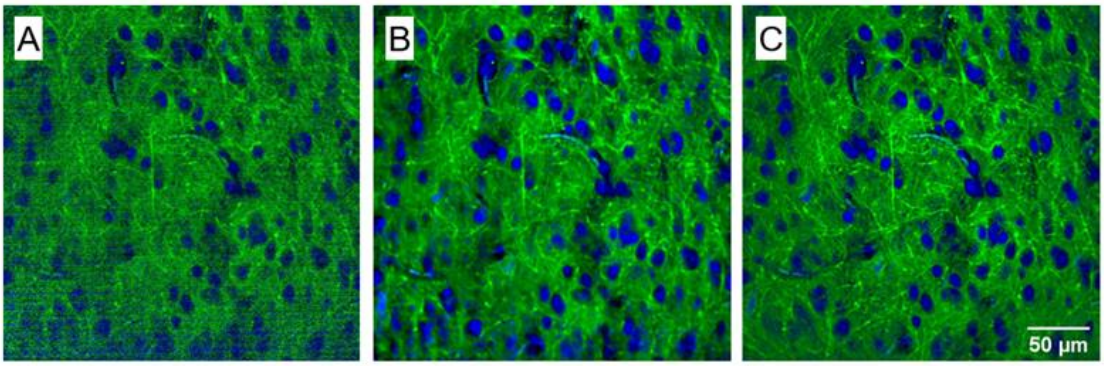

Figure 2.7. Denoising epi SRS images using the transmission-trained model. Two-color (lipids-green, proteins-blue) SRS images of coronal mouse-brain slice. Images (A-C) show an area of coronal mouse brain imaged in epi mode (A), the epi field of view denoised via deep learning algorithm (B), and the same field of view imaged in transmission mode (C).

Fig. 2.7A demonstrates how the noise starts to obfuscate the signal features even at high powers near the surface of the tissue in epi-SRS. In applying this algorithm to the epi SRS images, similar to the demonstrations shown in Figs. 2.1-2.6, image quality is significantly improved, approaching the quality of simultaneously acquired transmission-SRS images (Fig. 2.7B-C). This suggests an exciting method of improving *in vivo* epi-SRS imaging, whereby the epi images can be improved and denoised by a pre-trained algorithm using *ex vivo* tissue. This would effectively increase the imaging depth and utility of SRS imaging *in vivo*.

2.4 CONCLUSIONS

SRS imaging is a powerful label-free imaging technique that provides chemical information at sub-micron resolution. When using SRS imaging to examine biological systems, however, some limitations become evident. Scattering and absorption in tissue attenuates signal which ultimately limits image quality and imaging depths. Additionally, tissue can also be damaged by high power lasers but lowering laser power conversely lowers the signal. Finally, the

requirement of an epi imaging modality in *in vivo* applications inherently reduces the collected SRS signal as backscattered light is weaker than forward propagating light in transmission modality. These limitations stem from the same basis problem: signal strength is limited in the collection of an image resulting in noise dominated images.

While many general denoising algorithms have been developed for removing Poisson shaped shot-noise (mostly in the context of fluorescence imaging), none are aptly suited for the extremely low SNR's observed in the limits of SRS imaging. The general denoising algorithms fail to recover the inherent quantitative information in SRS images and often blur the relevant biological features.

Here we demonstrate the first use of deep learning to denoise and improve the quality of SRS images that outperforms more general denoising methods. Given the appropriate training, the deep learning denoising algorithms demonstrated here spatially recover relevant biological features (e.g. lipid droplets, axons, nuclei, etc.) without blurring or overfitting of features. Additionally, the deep learning algorithms appear to recover appropriate pixel values for truth images indicating the potential for recovery of quantitative information (pixel value plots from Figs. 2.2, 2.4, and 2.5). The main limitation of using a deep learning algorithm is the necessitation of acquiring an appropriate training set and the inherent trade off in generalizability of denoising in comparison to other denoising algorithms such as the here shown VST, PURE-LET, and CANDLE methods. For example, denoising the HeLa images with the mouse brain-trained algorithm or vice-versa exhibits worse performance than the appropriately trained algorithm for the given system. This trade-off, however, does not detract from the overall performance of the deep learning trained algorithm when used appropriately, especially considering the relatively small training sets used in these experiments (30 images for training). It is also worth considering

the relative generalizability of the trained algorithm of the system. For example, we demonstrated that training a single 30 image dataset is applicable to images acquired at different zoom, imaging depth, imaging power, and even imaging geometry. This demonstrates that the U-Net based algorithm created here should be treated as a specialized tool for improving SRS imaging. Additionally, while all results shown here are acquired with *ex vivo* samples, there is strong indication that *in vivo* images could be similarly improved given appropriate algorithm training (such as that shown in Fig. 2.7).

Ultimately, deep learning is valuable in augmenting the capabilities of SRS imaging in biological systems. Specifically, deep learning can improve the depths at which native biological information may be recovered, *in vivo* tissue imaging, and imaging of biomolecules at low abundance. The generalizability of this deep-learning based denoising approach may be improved in future work aimed at training a CNN on the noise profile in low-power SRS images directly. Additionally, utilizing a structural similarity (SSIM) loss function rather than an MSE loss function may provide more robust results.[93] That will be the subject of future studies. We expect future deep learning work in SRS imaging to expound on these improvements towards improving the utility of SRS imaging.

Chapter 3. EXTENDING THE IMAGING DEPTH LIMIT OF STIMULATED RAMAN SCATTERING MICROSCOPY

The work presented in this chapter has been published in the following article[94]:

A. H. Hill*, B. Manifold*, and D. Fu, "Tissue imaging depth limit of stimulated Raman scattering microscopy," Biomed. Opt. Express 11(2), 762 (2020).

*Denotes Co-first Authors

3.1 INTRODUCTION

Coherent Raman microscopy is severely limited in penetration depth for tissue imaging applications, similar to many other nonlinear optical imaging techniques such as two-photon fluorescence (TPF) and second-harmonic generation microscopy. Understanding the limitations of coherent Raman microscopy in tissue imaging is critically important to evaluate its potential for a wide range of ex vivo (e.g. intraoperative tumor margin detection), in vitro (e.g. drug screening), and in vivo (e.g. early cancer diagnosis) applications. Due to light scattering, it is widely recognized that the signal to noise ratio of coherent Raman microscopy deteriorates rapidly with imaging depth. Compared to TPF microscopy, imaging depth of coherent Raman microscopy is typically much less due to inherently much lower Raman cross-section of molecules. This limitation equally applies to both major variants of coherent Raman microscopy: stimulated Raman scattering (SRS) microscopy and coherent anti-stokes Raman scattering (CARS) microscopy. Both experiments utilize ultrafast laser pulses to coherently excite vibrational transitions in target molecules. Though their origin of signal and detection methods differ, both experiments typically operate in either transmission mode (for thin samples) or epi mode (for thick samples).

Unfortunately, to date there has been scarce report on the penetration depth limit of coherent Raman microscopy. Those studies which explicitly report the achieved imaging depths of coherent Raman microscopy in tissue suggest the limit to be around 20 μm – 100 μm[68,95–98], with varying tissue types being the cause of the large variance. Additionally, there have been no reports (to our knowledge) directly comparing the performance of epi and transmissive coherent Raman imaging in a side-by-side manner. Transmissive imaging is often advantageous in the case of optically thin samples and epi imaging is the default for optically thick samples (typically live animals or whole excised tissue). For samples of intermediate optical thickness (for example, sectioned tissues) there is no clear understanding of which imaging modality will provide the best results. A better understanding of these fundamental imaging limits and questions surrounding coherent Raman microscopy is crucial to the continuing development of these techniques as they continue gaining popularity in the biomedical imaging space.

Here we report a comprehensive study of epi and transmissive SRS signal size and imaging depth using a variety of ex vivo murine tissues (brain, lung, liver, and kidney). In this work we begin by characterizing SRS signal size and imaging depth through simultaneous epi and transmissive imaging of murine brain samples of varying thicknesses. As one would expect, epi (transmissive) SRS signal size and imaging depth increases (decreases) as tissue thickness grows, reflecting an increase (reduction) in backscattered (transmitted) photons. When tissue thicknesses reach 2 mm epi and transmissive signal sizes are roughly equivalent, and epi images provide slightly higher penetration depths than transmissive images. Using this method, we further characterized three additional tissue types (kidney, lung, and liver), each of which displays different scattering character reflecting their disparate structures and chemical compositions. Finally, we applied a recently developed convolutional neural network (CNN) based denoising

algorithm to SRS images and demonstrated the ability to resolve structural features at depths exceeding 210 μm in epi-imaging of brain, representing a 40 μm increase in imaging depth. We hope that this work will serve as a useful benchmark for the growing number of experimentalists entering the SRS microscopy field and a useful source to consult when considering how best to image a given sample.

3.2 METHODS

3.2.1 *Sample Preparation*

Murine tissue was harvested from recently sacrificed animals provided by UW Animal Use Training Services (AUTS) according to IACUC protocol 3388–03. After excision, samples of varying thicknesses were cut using razor blades with pre-measured spacers (250 μm, 500 μm, 1 mm, and 2 mm) to ensure accurate and uniform sample thicknesses. In the case of brain imaging, samples were prepared and imaged at all four thicknesses. Liver, lung, and kidney samples were all imaged at 1 mm thicknesses.

3.2.2 *SRS Imaging*

SRS images were collected using the homebuilt SRS microscope shown in Figure 2.1. The pump (800 nm) and Stokes (1040 nm) pulses are provided by the tunable and static outputs, respectively, of a dual output ultrafast oscillator (Insight DeepSee +, SpectraPhysics). The Stokes pulse is directed through an electrooptic modulator which modulates the 80MHz output pulse train to 20 MHz and provides the reference frequency used for lock-in detection. After modulation the Stokes pulse is sent onto a delay stage (used to control temporal delay between the pump and Stokes pulses) and then to a grating stretcher used to impart linear chirp[82]. The delay stage is used to determine the vibrational frequency images are collected at, and was positioned at the

temporal delay corresponding to 2920 cm^{-1} for the duration of these experiments. The pump pulse is directed through 60 cm of H-ZF52A glass to provide similar chirp. The pump and Stokes pulses are recombined at a dichroic mirror and routed to a set of galvanometer mirrors and finally into the back aperture of a microscope objective (Olympus XLPLN25XWMP2). Transmitted probe light is collected by a condenser lens and directed to a photodiode for lock-in detection. In the epi direction, backscattered and depolarized light is recollected by the focusing objective and isolated using a polarizing beam splitter. After filtering out residual Stokes light, backscattered pump photons are sent to a separate photodiode for lock-in detection. Transmissive and epi images were collected simultaneously for all samples and fields of view. SRS images were collected using 40 mW average power in both the pump and Stokes pulse trains. Each image samples a 285 μm × 285 μm field of view. All images were collected with an acquisition time of 4 s.

3.2.3 *Deep learning training and denoising*

Deep learning denoising was performed as reported previously[27]. Briefly, SRS images of murine brain samples were acquired at a depth of ~20 μm at low power (40 mW pump, 2 mW Stokes) and high power (40 mW pump, 40 mW Stokes) A U-Net deep learning architecture (publicly available code originally developed by Ounkomol et al.[77] and optimized for this application) was used to train a denoising algorithm that takes the low power (low SNR) images as input and predicts a corresponding high power (higher SNR) image of the same field of view similar to previously reported methods[99,100]. We then use the trained algorithm to denoise images taken deep in the murine brain at high power where SNR is low due to scattering and absorption of light in tissue.

The algorithm used here was supplied with 40 fields of view at both low and high power corresponding to signal and truth respectively. The 40 fields of view were randomly split into

10/30 test/train pairs. The algorithm was then trained over the course of 50,000 epochs using a learning rate of 0.001 with an Adam optimizer, momentum values of 0.5 and 0.999, and a batch size of 30 images. Following training, the algorithm was then fed the fields of view acquired deep in the brain for denoising. Training and denoising were performed on the University of Washington Hyak Mox supercomputer equipped with an Nvidia P100 graphics processing unit. The training session lasted ~10 hours, while denoising of the test and deep images took ~0.1 seconds per field of view.

3.3 Results and Discussion

3.3.1 *Effects of Tissue Thickness on SRS Signal and Imaging Depth*

As described above, murine brain tissue samples were cut to varying thicknesses (250 µm, 500 µm, 1 mm, and 2 mm) and simultaneously imaged in transmissive and epi geometries. Throughout each experiment, images were collected down to 240 µm at 5 µm intervals. The DC (unmodulated) and AC (modulated) portions of the epi and transmissive signal were recorded as a function of depth. Each measurement was performed in triplicate (on different fields of view) for each tissue thickness. Samples were prepared to enable imaging of areas with similar structural features and chemical composition across multiple samples. All brain data discussed in this manuscript refers to images collected of the cortex.

Transmissive DC signal size (in the form of absolute photocurrent detected) as a function of depth and tissue thickness is shown in Figure 3.1A. As one may expect, this signal represents total pump photons detected and the signal size is highest for the thinnest sample (250 µm, black) and decreases with thickness. Depth-dependent epi DC signal sizes, which increase as a function of tissue thickness, are shown in Figure 3.1B. The trends observed agree with what one might expect from thicker tissues enabling more scattering. Differences in the magnitude of the DC signal

between Transmissive and epi imaging largely reflect the differing collection efficiencies of the two imaging modalities.

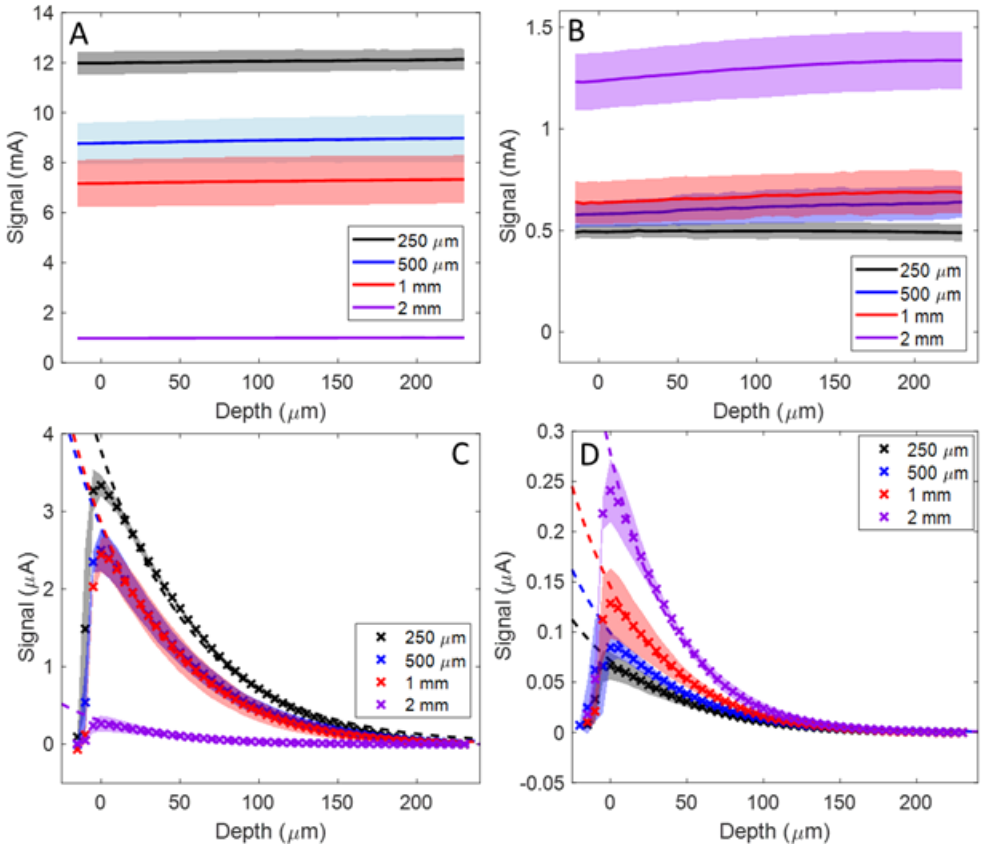

Figure 3.1. DC and SRS signal sizes for epi and transmissive images as a function of brain tissue thickness

(A) Transmissive DC signal size as a function of depth for four different tissue thicknesses. (B) Epi DC signal size as a function of depth for four different tissue thicknesses. (C) Transmissive SRS signal sizes as a function of depth for four different tissue thicknesses. (D) Epi SRS signal sizes as a function of depth for four different tissue thicknesses. X's and lines correspond to three-trial averages while shaded areas correspond to standard deviations.

Maximum recorded AC (SRS) signal sizes follow the same trends as the DC signal sizes for both transmissive and epi imaging. As expected, thinner samples (< 1 mm) yield significantly higher signals in transmissive mode than epi mode. When samples reach 2 mm thick, however, epi

and transmissive signal sizes are functionally identical. Maximum DC and SRS signal sizes in both epi and transmissive mode are compiled below in Table 3.1, along with the calculated modulation depths (the ratio of AC to DC signal) recorded using each imaging modality across all samples. In principle, the modulation depths recorded using epi and transmissive images should be identical. However, here we report roughly a factor of two shallower modulation depth for epi images in thin samples than those recorded in a transmissive geometry. This is possibly due to the fact that in thin tissue, a larger fraction of detected photons are due to reflection from the cover slide or tissue surface.

Table 3.1. Maximum DC and SRS Signal Sizes as a Function of Tissue Thickness and SRS Imaging Geometry in Murine Brain Tissue

Thickness	***DC Signal, Trans.* (mA)**	***DC Signal, Epi* (mA)**	***Max SRS Signal, Trans.* (μA)**	***Max SRS Signal, Epi* (μA)**	***SRS Modulation Depth, Trans.***	***SRS Modulation Depth, Epi***
250 μm	12.06 ± 0.44	0.50 ± 0.04	3.33 ± 0.16	0.07 ± 0.01	1.42×10^{-4}	6.0×10^{-5}
500 μm	8.89 ± 0.89	0.61 ± 0.08	2.5 ± 0.28	0.09 ± 0.02	1.44×10^{-4}	6.89×10^{-5}
1 mm	7.26 ± 0.97	0.67 ± 0.11	2.44 ± 0.23	0.13 ± 0.04	1.60×10^{-4}	9.55×10^{-5}
2 mm	0.99 ± 0.01	1.30 ± 0.14	0.26 ± 0.10	0.24 ± 0.03	1.10×10^{-4}	9.23×10^{-5}

Figures 3.1C and 3.1D show the recovered transmissive and epi SRS signal sizes as a function of depth, respectively. Unlike the DC signals, SRS signal decreases exponentially as a function of depth. The exponential decay of each curve can be fit to Equation 3.1 to determine tissue scattering length:

$$I_{SRS} = Ae^{-2z/L_s} \tag{3.1}$$

where A is maximum signal intensity, z is the depth beneath the surface, and Ls is the effective scattering length[101]. Scattering lengths recovered from 250 μm, 500 μm, 1 mm, and 2 mm using transmissive (epi) imaging are 119.3 ± 5.4 (104.2 ± 1.4), 110.3 ± 1.5 (101.9 ± 1.2), 105.0 ± 1.1

(95.7 ± 1.1), and 91.4 ± 1.9 (83.8 ± 1.1), respectively. We note that the exact scattering lengths generally match the lower end of previously reported range of 90 μm – 120 μm[102,103]. One likely contribution of low value measured here is the fact that unlike two-photon fluorescence, the use of two different wavelengths in SRS makes it more susceptible to aberration (both chromatic and spherical) induced signal degradation. This additional signal decrease with depth manifests as shorter scattering length. Detailed comparison with literature is further complicated by the dependence of scattering length on animal age and sample preparation.

3.3.2 *Imaging depth limit of SRS microscopy in Murine Brain*

While direct signal size measurements and comparisons are useful from a benchmarking point of view, the metrics of merit in many biological imaging experiments are signal-to-noise ratio (SNR) and effective imaging depth. For signal to noise measurements the signal value used was calculated as the mean pixel value of the field of view at a given depth. We chose this approach, as opposed to using the brightest features in an image to determine signal size, in an attempt to quantify the depth at which meaningful structural information could be observed. The standard deviation used in the calculation was the standard deviation of the deepest frame of each image stack. This frame was chosen since the negligible amount of SRS signal present at extreme depths suggests that the standard deviation would be dominated by shot noise as opposed to heterogeneities in signal magnitude across the field of view, as can be seen in images taken from shallower depths.

Figure 3.2 shows plots of epi and transmissive image SNR as a function of depth for four different tissue thicknesses. Maximum SNR for 250 μm thin tissue is 10 times higher in transmissive imaging compared with epi-imaging. The achievable imaging depths for transmissive and epi modalities (defined by the depth at which SNR first falls below 2.0, shown by black dashed

lines in Figure 3.2) are 205 μm and 110 μm, respectively. The difference becomes much smaller as tissue thickness increases. In agreement with the observed trends for both transmissive AC and DC signal size, maximum SNR and imaging depth decrease with tissue thickness for transmissive measurements. The maximum SNR for 2 mm thick tissue is 4 times lower. The SNR and imaging depth for the epi geometry follow the same trends seen for epi AC and DC signal sizes as well, in that they increase as a function of tissue thickness. 2 mm thick tissue yields maximum SNR that is 3 times higher than that in 250 μm thick tissue. Peak SNR and maximum imaging depth values for all four tissue thicknesses are compiled in Table 3.2.

In the case of thinner samples (250 μm – 1 mm) it is unsurprising that peak SNR is so much lower in epi images than transmission images. Since SRS microscopy is in general a shot-noise limited experiment, signal will scale as the square root of the average power of the detected pump pulse[85]. Based solely on the differences between transmission and epi DC signal sizes, thinner tissues should see a factor of 3.5 – 5 lower SNR in epi imaging as seen in Table 3.2. Here we report peak epi SNRs which are a factor of 3.7 – 9.7 lower than their transmission counterparts. This outsized loss of SNR in epi images can be traced back to the differences between epi and transmissive SRS signal modulation depths reported in Table 3.1.

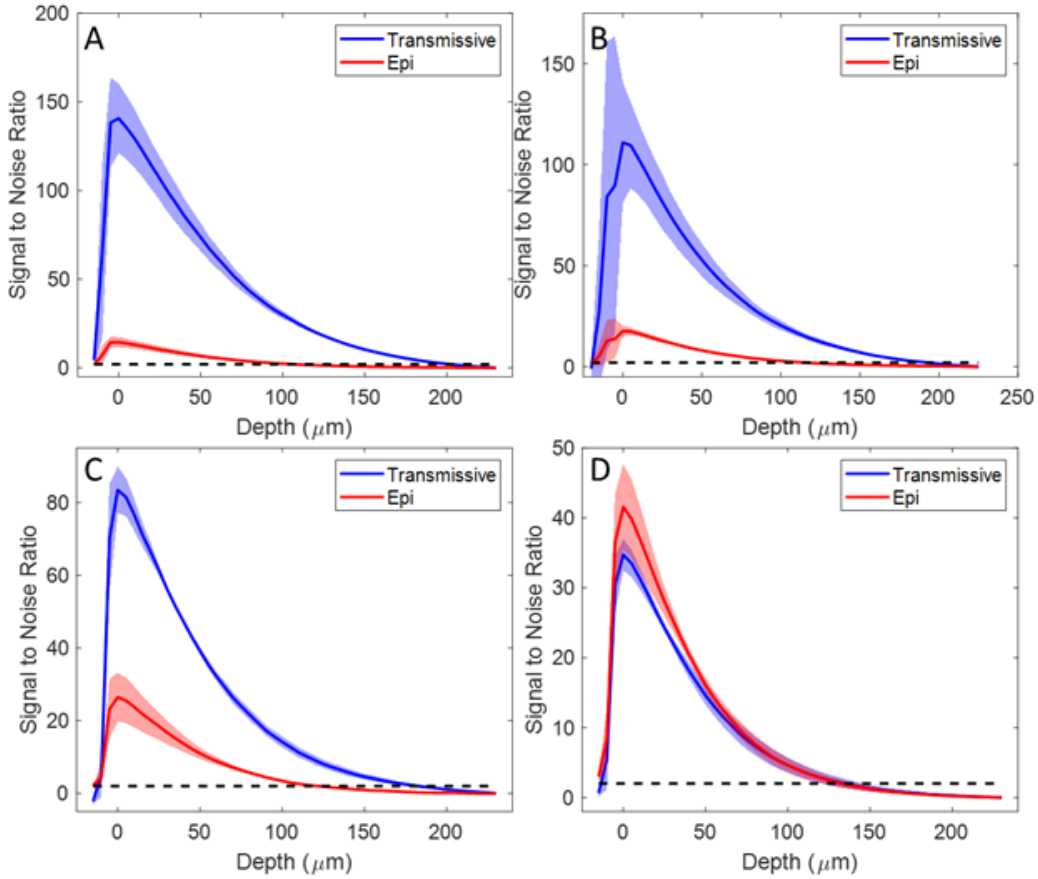

Figure 3.2. SNR as a Function of Depth for Murine Brain Epi and Transmissive Mode SRS Images in Samples of Varying Thickness

(A) SNR as a function of depth for a 250 µm thick tissue sample. (B) SNR as a function of depth for a 500 µm thick tissue sample. (C) SNR as a function of depth for a 1 mm thick sample. (D) SNR as a function of depth for a 2 mm thick sample. All curves shown are averages over three fields of view. Shaded areas reflect standard deviation within those datasets.

At a thickness of 2 mm, however, epi imaging yields a higher peak SNR (41.75 compared to 34.8) and a slightly deeper imaging depth (135 µm compared to 130 µm) than transmissive imaging. This reversal of SNR suggests that for brain samples thicker than 2mm, it is advantageous to use epi imaging. We expect that imaging SNR and penetration depth will further increase in intact brain due to even higher numbers of backscattered photons.

Table 3.2. Epi and Transmissive SRS Image Peak SNR and Imaging Depth for Murine Brain Tissue Samples of Varying Thickness

Thickness	***Peak SRS SNR, Transmissive***	***Achieved Imaging Depth, Trans. (μm)***	***Peak SRS SNR, Epi***	***Achieved Imaging Depth, Epi (μm)***
250 μm	140.6 ± 19.7	205	14.5 ± 2.9	110
500 μm	111.0 ± 29.0	195	17.6 ± 2.7	120
1 mm	100.5 ± 34.5	185	27.6 ± 5.9	125
2 mm	34.8 ± 2.3	130	41.8 ± 6.2	135

3.3.3 *Epi and Transmissive Imaging through Different Murine Tissues*

Following analysis of epi and transmissive SRS signals in murine brain tissue, we sought to gain a better understanding of SRS imaging in additional types of murine tissue. To that end, simultaneous transmissive and epi imaging experiments were conducted on 1 mm thick slices of murine lung, liver, and kidney tissue. Raw SRS and DC signal sizes in epi and transmissive modes for each tissue sample are shown in Figures 3.3A and 3.3B, respectively. When compared to the measurements conducted on 1 mm thick murine brain tissue slices, we can see that kidney and liver tissue yield rather similar transmissive DC signals but roughly half epi DC signals. Lung tissue, on the other hand, yields a transmissive DC ~60% lower and an epi DC signal four to eight times higher than all other studied tissues, suggesting that lung tissue is a much stronger scatterer than brain, kidney, or liver tissue.

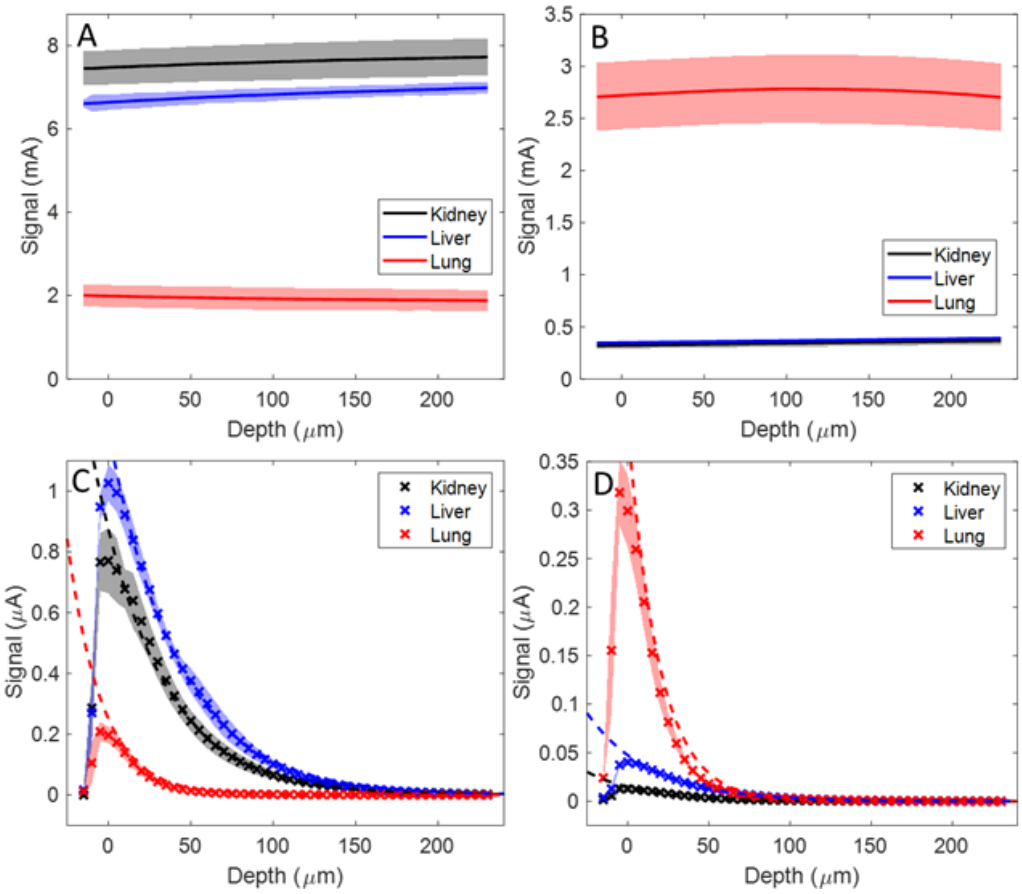

Figure 3.3. Transmissive and Epi SRS Signal Sizes for Different Murine Tissues

(A) Transmissive DC signal size as a function of depth in murine kidney, lung, and liver tissue. (B) Epi DC signal size as a function of depth in murine kidney, liver, and lung tissue. Dashed lines reflect fitting to equation 1. (C) Transmissive SRS signal size as a function of depth in murine kidney, lung, and liver tissue. (D) Epi SRS signal size as a function of depth in murine kidney, liver, and lung tissue. Dashed lines reflect fitting to Equation 1.

SRS signal sizes as a function of depth for each of the three tissue types during transmissive and epi imaging are shown in Figures 3.3C and 3.3D, respectively. To compare the scattering properties of each tissue type, the signal decay curves were fit with Equation 1 to determine the scattering length of photons in each tissue type. The results of the fittings are shown in Table 3.3. SNRs as a function of depth for each tissue type were also calculated for each of the three tissue types. The imaging depths and peak SNRs are tabulated in Table 3.3. For both kidney and liver

tissues, transmissive imaging provided a higher peak SNR than epi imaging, in agreement with the trends we report above for brain tissue. Transmissive imaging also achieved a significantly deeper imaging depth relative to epi imaging for kidney and liver tissues. In the case of lung tissue, however, epi imaging yielded a higher SNR and comparable imaging depth.

Table 3.3. Epi and Transmissive SRS Image Peak SNR and Imaging Depth for Various Murine Tissues

Tissue	L_s, Transmission (μm)	L_s, Epi (μm)	Peak SRS SNR, Transmissive	Peak SRS SNR, Epi	Achieved Imaging Depth, Trans. (μm)	Achieved Imaging Depth, Epi (μm)
Kidney	78.8 ± 0.9	70.5 ± 1.2	110.1 ± 11.5	10.1 ± 1.4	165	65
Liver	84.2 ± 1.7	77.1 ±0.7	136.3 ± 39.2	13.4 ± 5.3	170	80
Lung	41.0 ± 0.8	38.48 ± 0.7	52.8 ± 4.8	71.2 ± 5.2	70	70
Brain	95.7 ± 1.1	105.0 ± 1.1	100.5 ± 34.5	27.6 ± 6.0	185	120

3.3.4 *Deep Learning to Enhance Imaging Depth*

After comparing the epi and transmission SRS imaging modalities across a variety of murine tissues, we decided to explore possible avenues to increase the potential imaging depth of SRS microscopy. To that end, we employed our previously reported CNN-based denoising technique to determine whether its denoising capabilities would translate into deeper achievable imaging depths. Figure 3.4 shows the application of a machine learning denoising algorithm to epi collected SRS images at various depths. Panels 5A – 5D show data from an epi imaging experiment imaging a 2 mm thick sample of murine brain tissue at 15 μm, 100 μm, 170 μm, and 210 μm depths respectively. Figures 3.4E – 3.4H show versions of the images shown in 3.4A – 3.4D that have been passed through a machine learning based denoising algorithm. While all images show some improvement over their noisy counterparts, the effect truly becomes significant when

comparing Figures 3.4D and 3.4H. In Figure 3.4D some distinct features such as nuclei and axons are visible, but the quality of the image as a whole is quite low. By comparison, Figure 3.4H shows significantly more identifiable features and a higher quality image.

With respect to quantitative measurements of quality at these depths, the image shown in Figure 3.4C has an SNR of 2.1, which means it represents the maximum imaging depth based on the criterion we outline above. For the denoised images, however, SNR as defined above is not a responsible metric to compare image quality as there is no field void of signal features post-denoising. Other commonly used metrics such as peak SNR, root mean squared error, Pearson's correlation coefficient, and structural similarity index are also not useful here as a reliable truth image is not available for comparison at this depth of imaging. As such, signal to background (SBR) is calculated for each image and compared. SBR was chosen as a comparison metric over SNR because the SNR of images ran through the CNN denoising algorithm were found to be invariant with depth. We suspect this is caused by the algorithm adjusting the average value of the input images to better match the images used in its training, resulting in all images having very similar SNRs after denoising, regardless of the SNR of the input image. To calculate SBR, an area of low signal and few features is selected for both the noisy and denoised image (for example, the dark area in the bottom center of Figures 3.4C – 3.4D and 3.4G – 3.4H). The average pixel value of this area is taken to be the background. Then 6 lines spanning the field of view are selected in the image (3 horizontal and 3 vertical, each set of 3 equally spaced from one another). The peak pixel value from each of these lines is taken as signal plus background. The background value is subtracted from the peak value and then divided by the background value to give SBR. The SBR for the image is the average value from the 6 lines sampled on the image. We have chosen to use an area average for the background as noisy images (like those in Figures A.B.1 – A.B.2, shown

in the Appendix B) have large variance along a sampled line and using the minimum pixel value along such a line as the “background” will often grossly inflate SBR values for an image.

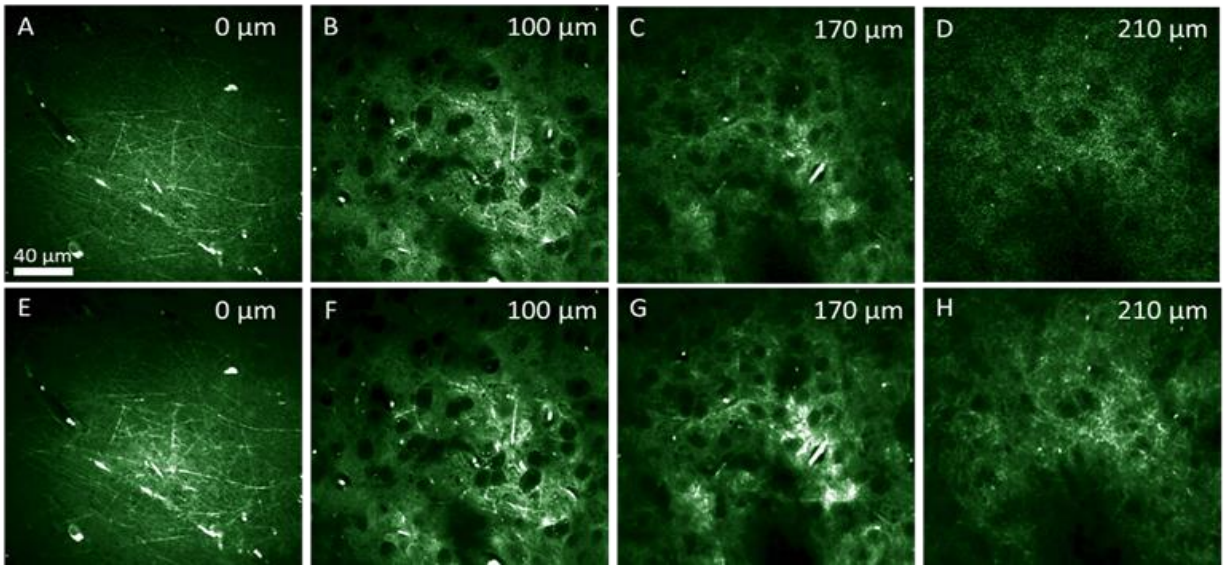

Figure 3.4. Demonstration of CNN-based denoising leading to an increased maximum imaging depth

(A) Epi SRS image collected at 0 µm depth. (B) Epi SRS image collected at 100 µm depth. (C) Epi SRS image collected at 165 µm depth, defined as the maximum imaging depth based on SNR. (D) Epi SRS image collected at 210 µm depth. (E) CNN denoised version of image in A. (F) CNN denoised version of image in B. (G) CNN denoised version of image in C. (H) CNN denoised version of image in D.

Using the outlined method for calculating the SBR, Figures 3.4C – 3.4D exhibit an SBR of 2.6 and 2.3 respectively. After denoising, the same fields of view shown in Figures 3.4G – 3.4H exhibit an SBR of 7.2 and 6.4 respectively. Thus, the deep learning denoising used here effectively improves the SBR of images at the defined limit of image quality by a factor of over 2.5. This data confirms that deep learning provides a promising avenue in extending the depth limitations of SRS microscopy. Line plots of the images used in this calculation are provided in Appendix B.

3.4 CONCLUSIONS

SRS microscopy is a powerful technique for the characterization of biological samples. When a transmissive collection is not possible, or the sample is thick enough that transmissive imaging results in minimal signal collection, an epi imaging geometry offers a viable alternative. The relative performances of epi and transmissive imaging are, as we demonstrate above, dependent on many parameters including tissue type and tissue thickness.

In the case of murine brain tissue, we report that tissue samples under 2 mm are most effectively imaged using a transmissive imaging geometry. Samples 2 mm or greater in thickness yield the best results in terms of peak SNR and imaging depth when imaged in an epi geometry. The threshold for which tissue thickness yields higher SNR in epi mode is strongly dependent on tissue scattering length. For example, lung tissue (which exhibited the shortest scattering length of all the interrogated tissues) yielded significantly better images in an epi geometry at a thickness of just 1 mm. We also report significant variation in absolute signal magnitude and achievable imaging depth between different tissue types, and even between samples of the same tissue type, confirming the role tissue heterogeneity plays in determining imaging quality.

Finally, we demonstrated the potential utility of CNN-based denoising algorithms towards achieving deeper maximum imaging depths. An algorithm developed for the purpose of these experiments was able to denoise images and increase imaging depth from 170 μm to 210 μm in 2 mm thick murine brain tissue samples. This 40 μm increase in imaging depth proves deep learning based denoising algorithms are poised to play a pivotal role in biological imaging in the coming years.

There are several other avenues to improve signal size and imaging depth in coherent Raman imaging that could be explored as well. In our measurements, the maximum pump photons

detected in the epi-direction is <10% of total pump photons. Saar et.al. has shown that using an annular detector [104] for epi imaging, up to 28% of photons can be collected. In combination with the polarizing beam splitter-based epi-imaging, it is possible to increase the collection efficiency by four-fold and sensitivity by two-fold, pushing imaging depth over 250 μm. Of course, this comes at the cost of additional experimental complexity[104]. Imaging at longer pump/Stokes wavelengths has also been shown to increase imaging depth in phantom samples due to increased scattering length, however this approach comes at the cost of SRS signal intensity[105]. The actual benefit in using long wavelength for tissue imaging warrants further study. Correcting for optical aberration is another approach that can push the imaging depth even lower. Previous report shows that coherent Raman signal can be increased by 6-fold in muscle tissue[95], potentially allowing another increase of over 50-100 μm. Tissue clearing methods can also be used to achieve deeper imaging depths[68,106], though some methods have been shown to alter the structure of the cleared tissue[107,108]. However, tissue clearing (with the exception of skull clearing[109]) is incompatible with live processes and therefore is limited to study of fixed tissue.

In conclusion, we have conducted a comparative study of epi and transmissive imaging in various types of murine tissue. Throughout this study we have characterized epi and transmissive imaging efficacy as a function of tissue thickness and tissue type, and we report the recovered scattering lengths, maximum imaging depths, SNRs, and absolute signal magnitudes. Additionally, we have shown that CNN-based denoising algorithms can increase the maximum imaging depth in coherent Raman microscopy experiments, though further experiments are required to ascertain the quantitative utility of this approach.

Chapter 4. THE U-WITHIN-U-NET: A VERSATILE DEEP LEARNING ARCHITECTURE FOR CLASSIFICATION AND LABEL-FREE PREDICTION OF HYPERSPECTRAL IMAGES

The work presented in this chapter has been published in the following article[110]:

Manifold, B.; Men, S.; Hu, R.; Fu, D. A Versatile Deep Learning Architecture for Classification and Label-Free Prediction of Hyperspectral Images. Nat Mach Intell 2021, 3 (4), 306–315. https://doi.org/10.1038/s42256-021-00309-y.

4.1 INTRODUCTION

Computer vision techniques based on deep learning have recently demonstrated a myriad of novel applications in many disciplines. With the continuous improvement and availability of advanced computing hardware and open-source methods, deep learning is finding broader use in a wide variety of imaging, sensing, and biophotonics research[25,26]. The flexibility of deep learning for image processing enables facile adoption of existing frameworks for many different imaging modalities such as transmitted light microscopy, fluorescence microscopy, X-ray imaging, magnetic resonance imaging, and many more[111–115]. Often the images from such techniques are passed to a deep learning algorithm to perform tasks like classifying diseases, segmenting spatial features, improving image quality, or predicting alternate imaging modalities[99,116–118]. However, the majority of work done so far performs deep learning on monospectral images. Such monospectral images contain only a single intensity value at each pixel. That is, there is no spectral information inherent to the imaging technique such as in black-and-white photography, X-ray imaging, or magnetic resonance imaging. Contrary to monospectral images are multispectral and hyperspectral images where multiple spectral components of a field

of view can be depicted in their own image. We take “multispectral” to be a subset of “hyperspectral“ specifically pertaining to images that contain relatively few spectral channels (e.g. RGB imaging). Hyperspectral imaging combines spectroscopy and imaging such that each pixel of the image contains a wide spectral profile that allows for detailed characterization.

Linear decomposition, phasor analysis, support vector machines and other machine learning methods have indeed been used for analysis of hyperspectral imaging datasets[12–16,119,120]. While many of these techniques have demonstrated promising results, such methods may suffer from limited generalizability or information loss, limiting their ultimate performance[121,122] Deep learning, in contrast, potentially offers a method for learning based on both spectral and spatial signatures and their nonlinear interplay allowing for improved performance in a variety of hyperspectral imaging analysis tasks[123,124]. However, techniques for these hyperspectral stacks face unique challenges in computer vision research[20,21]. For example, standard deep learning architectures that work for monospectral images (consisting of 2 or 3 spatial dimensions), may not work for hyperspectral stacks due to the extra dimension needed for spectral information. Frameworks such as Mayerich et al’s Stain-less Staining[125] or Behrmann et al’s work in mass spec imaging[126] address this by interpreting the spectra at individual pixels of hyperspectral images to produce excellent results in label-free prediction and classification, but may be missing contextual information from spatial convolutions of the whole image. Zhang et al’s recently published work bypasses the need for spectral deep learning by using machine learning to interpret spectral information and create truth maps to which spatial deep learning of images can be trained[127]. Other frameworks for hyperspectral deep learning based on spectral-spatial convolutions also exist but are often rigid; only performing a particular task like binary pixel or multi-class label classification[22–24]. Further, a convolutional framework for

predicting entirely alternate imaging modalities (where the final number of spectral channels is unlikely to match the input, but spatial resolution is maintained) from hyperspectral images, to our knowledge, has not been reported. We thus present a new architecture, the U-wthin-U-Net (UwU-Net) to address these current shortcomings in hyperspectral deep learning and improve the utility of hyperspectral imaging techniques.

The UwU-Net Architecture presented here is based on the U-Net architecture developed originally by Ronneberger et al where a specialized autoencoder encodes and decodes spatial feature information in an input image to reconstruct some new output image[84]. The U-Net separates itself from a traditional autoencoder with the recontextualization of information through concatenations at equivalent encode-decode levels (noted as blue arrows in Figure 4.1a). This eliminates the discarding of information as in a traditional autoencoder. While the original work was concerned with image segmentation, the U-Net has seen use in a variety of applications including segmentation, label-free prediction, and denoising[27,99,128,129]. However, most works that utilize the U-Net in this way are not concerned with images that contain multiple spectral channels. Indeed, the original U-Net is generally not applicable to hyperspectral images as the architecture is dedicated to encoding multiple spatial feature channels starting from a single spatial channel image as shown in Figure 4.1a. The typical 2D kernel of a U-Net is thus not well suited for hyperspectral stacks which have a third tensor dimension dedicated to spectral channels. A 3D kernel could potentially be used, but then the spatial and spectral information are being mixed during the feature encoding in a problematic fashion for image reconstruction[130]. Modification of input and output layers to match spectral dimensions is often useful in the multispectral regime but may be too facile of a change to adequately handle spectrally complex hyperspectral images. While other recently reported modifications to the U-Net have also shown

improvements with respect to the original U-Net on semantic segmentation and classification of remote sensing datasets (some of which involve multispectral datasets)[131–133], we report a robust architecture for multiple hyperspectral imaging tasks.

To create a hyperspectral deep learning architecture with the robustness and features of the traditional U-Net, we have amended the U-Net architecture such that spectral channel information is handled by a separate “U” structure “outside” of an arbitrary number of traditional spatial U-Nets as shown in Figure 4.1b. This UwU-Net architecture allows dedication of tunable free parameters to both spectral information (outer U) and spatial information (inner U’s). The architecture’s parameters can be empirically tuned to change the spectral layer depth, number of spatial U’s at the center, or output spectral size based on the dataset. Here we demonstrate the utility of this new architecture in 3 different tasks on 3 different types of hyperspectral imaging: feature segmentation and classification on the high altitude hyperspectral imaging Indian Pines dataset, monoisotopic drug location prediction in rat liver from mass spectrometry images, and label-free prediction of cellular organelle fluorescence in stimulated Raman scattering (SRS) microscopy.

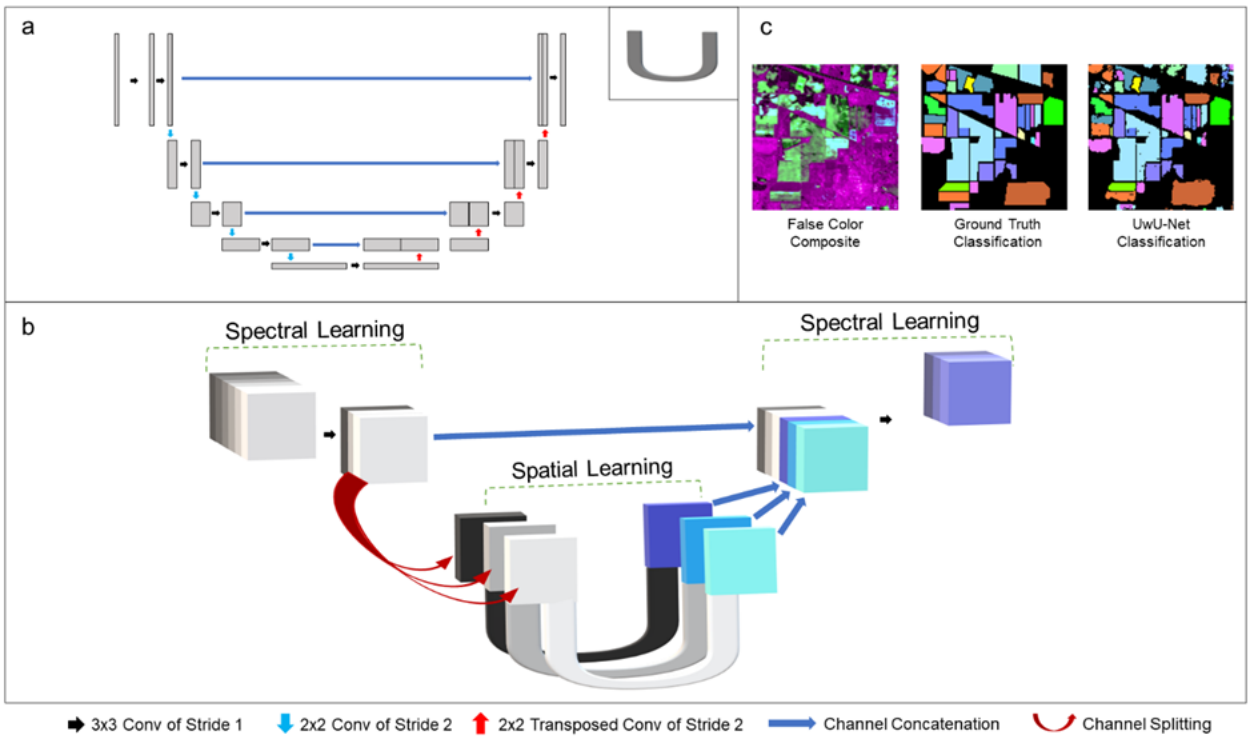

Figure 4.1. Architecture Diagrams and Indian Pines Classification

Panel a shows a schematic representation of the traditional U-Net (adapted from Ounkomel et al.[129]) where a single 2D image is convolved to encode and decode spatial features. The "U" in the upper right corner of panel a denotes its schematic representation as used in panel b. Panel b shows the schematic representation of the UwU-Net where an arbitrarily dimensioned hyperspectral stack is convolved both spectrally and spatially to produce an arbitrarily dimensioned output stack. The symbols used in panels a and b are noted at the bottom of the figure to show their operational meanings. Here, "conv" is short for convolution and the "NxN" shown describes with pixel size of the kernel used for convolution. Panel c depicts a false color composite of 3 different spectral bands from the original 200-band hyperspectral stack, the truth classifications, and predicted classifications from the UwU-net.

The first task concerns segmentation and classification of the Indian Pines dataset which depicts a scene of farmland in northwest Indiana across a large range of wavelengths spanning the ultraviolet to short infrared region (400-2500 nm)[134]. The publicly available dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer and provides a model task for

hyperspectral deep learning: segmentation and classification of various crop and foliage types. The broad spectrum and spatial heterogeneity of the scene demonstrates a deep learning algorithm's ability to correctly identify and segment features based on both spectral signatures and spatial positions. Moreover, the use of this dataset by previous work in hyperspectral deep learning allows for comparison of our proposed architecture[135–137].

The second task concerns predicting drug location in a model rat liver tissue sample from mass spectrometry imaging. Mass spectrometry imaging (MSI) is a powerful technique that provides spatially resolved, highly specific chemical information in the form of molecular ion masses. Where most deep learning computer vision work is centered around interpretation of optical images, MSI is particularly interesting to approach with deep-learning as it has an enormous spectral dimension that provides highly specific, but difficult to interpret *in situ* chemical information[138,139]. Most MSI work follows from traditional linear decomposition and analysis that is well-developed and ubiquitous in mass spectrometry[140–144]. Deep learning has been demonstrated for MSI datasets[126,138,145,146], but has been chiefly used for spectral dimensionality reduction or interpretation. To our knowledge, the simultaneous interpretation of spatial and spectral information using convolutional deep learning in MSI has yet to be reported. We demonstrate one way the UwU-Net architecture could be used in MSI by simultaneously predicting the highly specific monoisotopic peak locations of 12 drugs from low mass resolution binned images.

Finally, the third task demonstrates the capability of the UwU-Net to perform label-free prediction of fluorescence images from SRS microscopy images. SRS microscopy is a hyperspectral imaging technique where molecular vibrational bonds are coherently interrogated by two ultrashort laser pulses[9,34,147]. While the vibrational information afforded by SRS

microscopy can be specific to a given molecule, there are often many overlapping contributions to vibrational signals that confound image interpretation. In this work, we show that the specificity of SRS microscopy can be improved by deep learning to predict fluorescence images that are highly specific to an organelle. Further, we show that the trained algorithms can be multiplexed to create label-free cell organelle images in live cells.

4.2 METHODS

4.2.1 *Cell Sample Preparation*

A549 cells were cultured in ATCC F-12K medium with 10% fetal bovine serum at 37 °C with 5% CO2 atmosphere. Cells were seeded on coverslips 24 hours prior to imaging. Fixed cells were first dyed then fixed using 2% paraformaldehyde. Live cells were first mounted, imaged with SRS and then stained for fluorescence imaging. The fluorescent dyes used were Hoescht 33342, MitoTracker Red CMXRos, and ER-Tracker Green for nucleus, mitochondria, and ER respectively. All dye protocols were based on the provided instructions from the manufacturer.

4.2.2 *Simultaneous SRS and 2-Photon Fluorescence Microscopy*

SRS Microscopy was performed on a homebuilt SRS microscope as described previously. Briefly, an Insight DeepSee+ provides synchronized 798 nm and 1040 nm laser pulses which are passed through high density glass and a grating stretcher pair, respectively, to control pulse chirp. The 1040 nm bean is modulated by an electro optical modulator and polarizing beam splitter to operate in the stimulated Raman loss scheme. Time delay of the 1040 beam was controlled by a computer-controlled Zaber X-DMQ12P-DE52-KX14A delay stage. Both pulses are combined on a dichroic mirror before being directed through the microscope by a pair of scanning galvo mirrors. The microscope is a Nikon Eclipse FN1 equipped with a 40x 1.15 NA objective. The 800 and 1040

nm laser powers were set to 20 mW at focus for both beams in all experiments. Light passed through the sample is collected by a 1.4 NA condenser, filtered by a 700 nm long pass filter (to remove fluorescence light) and 1000 nm short pass filter (to remove the 1040 nm light), and finally collected on a homebuilt photodiode connected to a Zurich Instruments HF2LI lock-in amplifier. Two photon fluorescence is captured in the backwards direction by a 650 long pass dichroic towards a photomultiplier tube. SRS signal from the lock-in amplifier and fluorescence signal from the photomultiplier tube were collected simultaneously using ScanImage[148]. Images were acquired with 512 x 512 pixels and a pixel dwell time of 8 μs at each of the 10 vibrational transitions as noted in Figure 4.3d. It is noted that the transitions noted in Figure 4.3d represent only the center of the probed band with 19 cm^{-1} spectral resolution. This means that at the step size of ~15 cm^{-1} per image in the stack, the full CH region is probed during hyperspectral imaging.

4.2.3 *UwU-Net Functional Description*

An input hyperspectral stack of dimensions (L, X, Y) is first passed to the architecture. Here, L represents the number of input channels of the hyperspectral stack (e.g. 200 for Indian Pines, 300 for MSI drug location prediction, or 10 for SRS images) and X and Y are the number of spatial pixels in the image (in all cases here X = Y). The stack is first reduced to (M, X, Y) in the channel dimension, where L > M, with a 3x3 kernel convolution of stride 1 over all L channels followed by a batch normalization and rectified linear unit (ReLU) activation function. The new stack is then reduced once more in the channel dimension by the same process to a stack of (N, X, Y) where N is the desired final number of spatial tuning channels. The stack is then split at the channel dimension (if N > 1) such that there are now N number of (X,Y) images. Each of these images is passed to its own U-Net for spatial feature learning as described previously[129]. The resulting N number of images from each spatial U-Net are then reconcatenated in the channel

dimension to reform a (N, X, Y) stack. This (N, X, Y) stack is then concatenated in the channel dimension to the (N, X, Y) stack from prior to splitting (mimicking the recovery of information as in the traditional U-Net) to form a stack of (2N, X, Y). This (2N, X, Y) stack is reduced to (O, X, Y) by a 3x3 kernel convolution of stride 1 over the 2N channels followed by a batch normalization and ReLU activation function. This predicted stack is then compared to the truth stack (also of dimension [O, X, Y]), a mean squared error is calculated for all channels, and parameters are tuned in a backpropagating fashion.

4.2.4 *Training Parameters, Data Preparation, and Computational Hardware*

The models trained and shown in this paper were developed and built using the pytorch-fnet framework originally developed by Ounkomol et al[129]. All models were trained using the pytorch-fnet default parameters with a few exceptions. The models were trained using randomized starting parameters on batches of randomized patches from the given dataset. Model parameters are tuned in a stochastic gradient descent manner based on minimization of mean squared error. The pytorch-fnet framework utilizes an Adam optimizer with a 0.001 learning rate and beta values of 0.5 and 0.999. The rat liver drug prediction model which was trained only for 23,000 iterations due to the satisfactory prediction accuracy and long training iteration time. The Indian Pines and rat liver drug prediction models were trained with buffer size of 6 due to the reduced number of training datasets. The Indian Pines classification and rat liver drug prediction model used patch sizes of 64 x 64 pixels for training, while all organelle prediction models utilized patch sizes of 256 x 256 pixels.

Nearly all image preparations and processing discussed below were performed using Fiji, an imageJ platform. The exception was the additional use of Datacube Explorer for initial processing of the raw MSI data.

The 200 band Indian Pines dataset was used natively from the published source. The ground truth stack was created by separating the individual labeled images via thresholding then concatenating all truth images into a TIF stack. The native pytorch-fnet cropper was used to crop the images to 144 x 144 pixels from 145 x 145 pixels to accommodate the spatial learning in the central U-Nets of the UwU-Net architectures. Training data was augmented by rotations and flips with the original dataset withheld for testing. This equated to 6 training datasets and 2 test datasets. Final predictions were recolored for each label and then overlaid into the shown prediction image (Figure 4.1c). The UwU-Nets reported in Table 1 use 1 (1-U) or 17 (17-U) spatial U-Nets at their center during training.

The rat liver MSI dataset was first prepared by saving the 330-630 m/z window at 1 m/z bins from the raw data using Datacube Explorer. All 300 images were concatenated into a TIF stack using Fiji. The monoisotopic images at 0.001 m/z resolution were then saved for each drug using Datacube Explorer following the m/z peaks and appropriate FWHM bins as noted by Eriksson et al[149]. The 12 drug peak images were concatenated into a TIF stack using Fiji. Both stacks were padded with zeros in Fiji from their native 247 x 181 pixel size to 256 x 256 pixels to be compatible with the spatial U-Nets within the UwU-Net architecture. Training data here was also augmented by rotations and flips with the original dataset withheld for testing. There were 7 datasets used for training. The shown 1 m/z bin, truth peak, and predicted peak images for the drugs were normalized, contrast adjusted to the same level, and colored using the “Red Hot” Fiji lookup table. The UwU-Nets reported in Table 4.1 use 1 (1-U), 5 (5-U), or 12 (12-U) spatial U-Nets at their center.

The simultaneously collected SRS and Fluorescence images were first separated into 2 respective TIF stacks. The SRS stack was used as is for training and prediction. The fluorescence

stacks were averaged to a single image and used as the truth for training and prediction. The fixed cell nucleus, mitochondria, and ER models utilized 43, 46, and 35 images, respectively, with a randomized 80%/20% train-test split for each model. Images predicted by the model were normalized, contrast adjusted to the same level, then colored using the “mpl-inferno”, “Cyan”, “Green”, and “Magenta” Fiji lookup tables for SRS, nucleus, mitochondria, and endoplasmic reticulum, respectively.

All model development, training, and prediction as well as image processing was performed on a homebuilt machine running Ubuntu 18.04. The machine is equipped with an AMD 2950X processor, Nvidia Titan RTX graphics processing unit, 64 GB memory, and a 2 TB solid state drive. All dependency software versions were based on the pytorch-fnet requirements. On our machine, trainings for Indian Pines, rat liver drug, and organelle models took ~4, ~8, and ~5 hours respectively. In all models, prediction of individual test images took 1 second or less.

4.2.5 *Quantitative Metrics*

Prediction quality was assessed by overall accuracy (OA), Intersection Over Union (IOU) Pearson’s correlation coefficient (PCC), normalized root mean squared error (NRMSE), and feature similarity index (FSIM)

OA is used to evaluate the binary pixel values assigned for each classification. Here, the number of errantly predicted pixels are counted, subtracted from the total number of pixels, then divided by the total number of pixels. A percentage score is reported here where accuracy closer to 100% indicates a more accurate prediction.

IOU also measures the segmentation and classification accuracy by taking the ratio of the intersection between predicted pixels and true pixels (i.e. true positives) and union of predicted pixels and true pixels (i.e. true positives plus false positives). The resulting ratio indicates how

accurately the model segments and classifies areas where values closer to 1 indicate more accurate prediction.

PCC is used to correlate the pixels of the truth and predicted images. The covariance of the two images is divided by the standard deviation of the two images to provide a value indicating pixel-to-pixel correlation. A PCC of 1 would indicate perfect correlation while 0 would indicate no correlation.

NRMSE is used to express the accuracy of a predicted pixel versus the same pixel in the truth image. Here a value closer to 0 indicates a more accurate prediction model.

FSIM is used as an image quality assessment metric that mimics human perception of image similarity. Like the structural similarity index (SSIM), FSIM incorporates the spatially associated pixels in the images during calculation to provide a better notion of perceived similarity. However FSIM emphasizes low-level features of images to more accurately reflect the human visual system's perception of image similarity[150]. Here an FSIM of 1 indicates perfectly similar images while 0 would indicate no similarity.

Quantitative metrics were calculated using Fiji "Coloc 2" (PCC), and "SNR" (NRMSE) plugins on the normalized images produced by the trained prediction model. FSIM was calculated using the MATLAB code provided by Zhang et al[150], following the prescribed instructions.

4.3 Results and Discussion

4.3.1 *Indian Pines Classification*

To demonstrate this flexibility and to validate the architecture's capability to classify an arbitrary number of features from hyperspectral images, a 1-U UwU-Net (where there is 1 spatial U-Net at the center of the architecture) and 17-U UwU-Net (where there are 17 spatial U-Nets at the center) were trained to classify the Indian Pines AVARIS dataset[134]. The hyperspectral

images consist of 200 spectral channels (where 20 of the original 220 bands have been removed due to water absorption) across a broad range of wavelengths (400-2500 nm) with 144 x 144 pixel images (cropped from 145 x 145 to be compatible with the spatial U-Nets) at each wavelength. The images contain a high-altitude 2 mile by 2 mile field of view of farmland in northwest Indiana. The ground truth images consist of non-mutually exclusive hand-drawn maps of the various crops and foliage depicted in the field of view. In total, there are 16 classifications shown in Figure 4.1c and listed in Table 4.1. Here, the UwU-Net is trained to predict a 17 x 144 x 144 image stack (16 classifications plus an unused background) from the 200 x 144 x 144 input image stack. The initial 200 channels are first reduced via convolution to 100 then to the final 1 (for the 1-U UwU-Net) or 17 (for the 17-U UwU-Net) before spatial learning. The output predicted images are thresholded to create a binary map to compare against the ground truth image.

Table 4.1. Classification Accuracy of the Indian Pines Dataset

Classification Label	**UwU-Net (1-U)**	**ResNet**[136]	**MPRN**[136]	**AU-Caps-Gan**[137]	**UwU-Net (17-U)**
Alfalfa	97.40	98.33	98.89	99.15	**99.96**
Corn (No Till)	93.66	99.28	**99.51**	99.50	98.57
Corn (Min Till)	95.98	98.80	98.92	99.12	**99.19**
Corn	98.83	98.20	98.52	98.34	**99.78**
Grass (Pasture)	97.60	97.97	97.92	98.70	**99.48**
Grass (Trees)	98.26	98.80	99.08	99.42	**99.80**
Grass (Mowed Pasture)	99.98	**100**	98.18	98.74	99.98
Hay (Windrowed)	97.65	**100**	**100**	99.27	99.91
Oats	99.90	97.50	97.50	98.68	**99.98**
Soybeans (No Till)	96.35	97.99	98.14	98.45	**99.18**
Soybeans (Min Till)	79.37	99.27	**99.38**	99.12	98.49
Soybeans (Clean Till)	97.15	98.35	98.69	98.34	**99.23**
Wheat	99.50	99.14	98.90	98.69	**99.93**
Woods	94.87	99.88	**99.98**	99.33	99.18
Buildings	98.07	99.55	**99.68**	99.41	99.18
Stone-Steel Tower	99.74	94.52	96.44	98.94	**99.91**
Overall Accuracy	96.52 ± 4.7	99.01	99.16	99.12 ± 0.25	**99.48 ± 0.50**

The individual and overall classification accuracy of the Indian Pines dataset from various hyperspectral deep learning models and the presented UwU-Net model. Note the ResNet, MPRN, and AU-Caps-GAN models are reported as produced in their respective references, the ResNet and MPRN classifications were reported without uncertainties. Reported uncertainties refer to the standard deviation among the n= 16 classifications.

Looking at the results in Table 4.1, the 17-U UwU-Net performs well with nearly all classifications exceeding 99% accuracy. The exceptions are the classification of an untilled corn field in the upper left of the field of view that are instead identified as a mixture of the three soybean classifications. We also note the prediction of crops at the top-middle, top-right, and bottom of the field of view. While these areas contribute to the error, we note that crops do exist in these parts of the hyperspectral images (as seen in the composite image in Figure 4.1c) but are unidentified in the hand-drawn truth maps. To better reflect the model's performance, especially in these cases, counts of false positive and negative pixels and the intersection over union (IOU) for each class is provided in Appendix C Table A.C.1. The overall accuracy (99.48% ± 0.50%), however, is in concordance with state-of-the-art architectures for hyperspectral classification on the Indian Pines dataset[136,137,151–153]. Three of these architectures' (ResNet, Multi-Path ResNet, and Auxillary Capsule GAN) classification accuracies are shown in Table 4.1 for comparison with the 17-U UwU-Net demonstrating the highest accuracy. We note that the 1-U UwU-Net (with its more modest modifications to the original U-Net) performs worse than the other models suggesting that the additional spatial parameters afforded by the parallel U-Nets at the center of the UwU-Net contribute towards a more accurate model. For additional comparison, a basic U-Net (where the initial and final layers have been simply adjusted to accommodate the desired input/output channel number) was also trained. However, it was unable to classify any of the labels properly suggesting that UwU-Nets spectral layers are critical for proper identifications. A representative example of

one of the basic U-Net's errant classifications is shown in Appendix C Figure A.C.1. These results demonstrate the UwU-Net's ability to simultaneously segment and classify features from hyperspectral images with high accuracy. However, the UwU-Net is not limited to a binary pixel classification, like some hyperspectral architectures here compared, but can also predict intensity features as shown in the demonstrations below.

4.3.2 *Drug Location Prediction in Mass Spectrometry Images*

To further demonstrate the utility of the UwU-Net in deep learning of hyperspectral images, we predict the location of multiple drugs (most of which are cancer treatment drugs) in a rat liver slice from publicly available mass spectrometry imaging data originally taken by Eriksson et al[149]. Here, a frozen-fixed rat liver section was spiked with 5 mixtures of diluted drugs, where each mixture contains some combination of 4 of the 12 potential drugs at varying concentrations. MSI was then performed on the liver slice in the mass-to-charge-ratio (m/z) range of 150 – 1000 m/z at a mass resolution of 0.001 m/z. This means that this particular raw hyperspectral dataset contains 850,000 images which is not uncommon for MSI datasets. Given this colossal spectral density, MSI datasets must be narrowed to small "windows" (e.g. only 1000 images between 300.000 m/z – 300.999 m/z are shown) and/or "binned" (e.g. all the 0.001 m/z images from 300.000 m/z – 300.999 m/z are summed together to form a single 1 m/z bin image) to be viewable. Both windowing and binning sacrifice information for interpretability. Windowing allows for only seeing a few mass components at a time while binning sacrifices the hallmark specificity of mass spectrometry[154]. Analysis of these large datasets can also be cumbersome, taking potentially hours or longer to interpret per dataset.

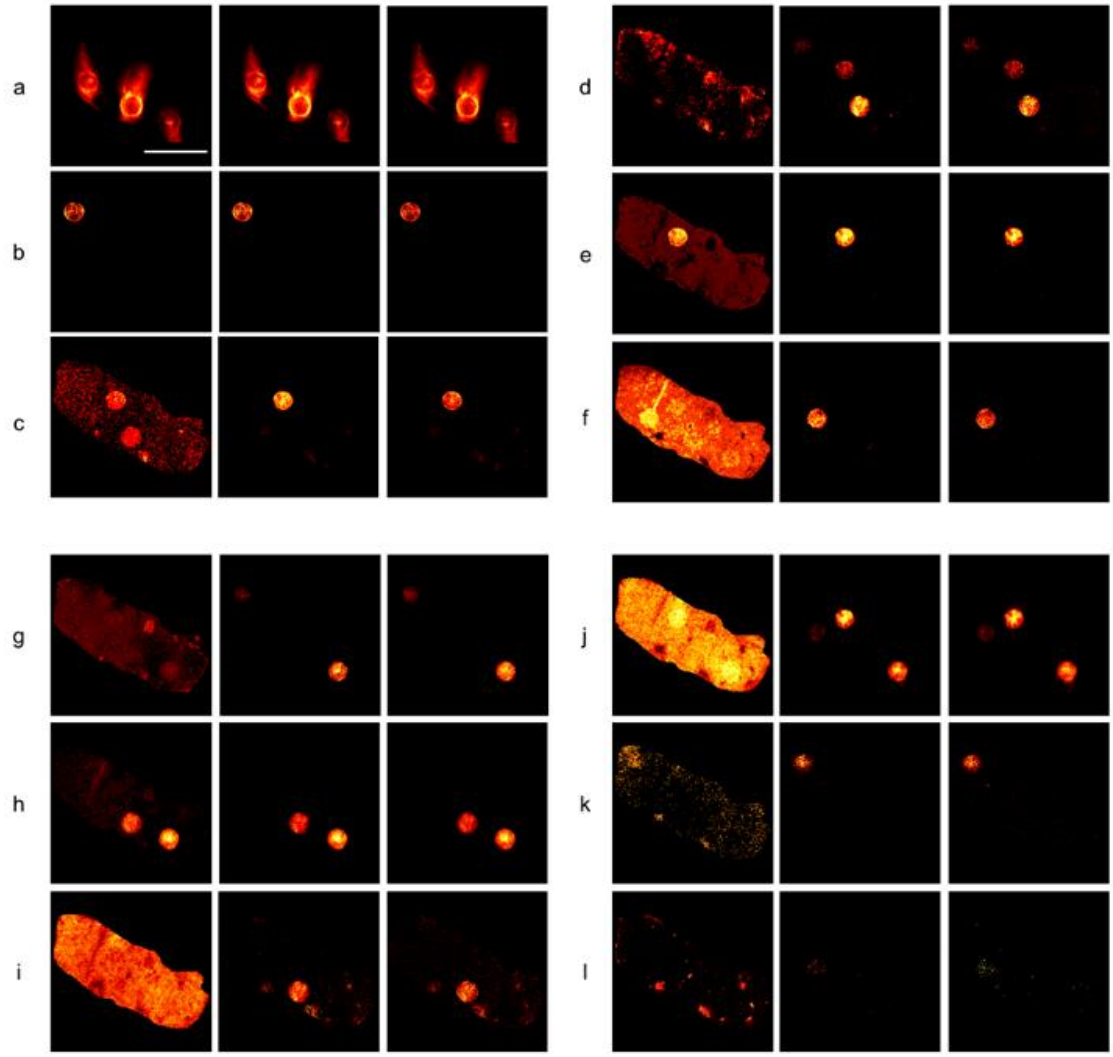

Figure 4.2. Mass spectrometry images of drug-spikes rat liver slice

Each row (a-l) shows (from left to right) a 1 m/z bin image from the input 300 image hyperspectral stack that contains a given drug's monoisotopic peak, the 5-U UwU-net predicted 0.001 m/z bin image of the drug, and the 0.001 m/z bin image specific to that drug's monoisotopic peak. The following drugs are depicted in their respective panels: Ipratropium (panel a), Vatalanib (panel b), Erlotinib (panel c), Sunitinib (panel d), Pazopanib (panel e), Gefitinib (panel f), Sorafenib (panel g), Dasatinib (panel h), Imatinib (panel i), Dabrafinib (panel j), Lapatinib (panel k), Trametinib (panel l). Scalebar = 4 mm.

The work we present here demonstrates a potential solution to this information trade-off issue by predicting high mass resolution drug location images (corresponding to each drug's

monoisotopic peak) from a window of hyperspectral low-resolution binned mass images of the spiked rat liver tissue. Specifically, the region of 330 – 630 m/z (a window containing all monoisotopic drug peaks) was binned into 1 m/z images and concatenated into a hyperspectral image stack. Then, the 0.001 m/z resolution images corresponding to the monoisotopic peaks of the 12 drugs (as determined in the previous publication) were isolated from the raw MSI data and concatenated to produce a stack where each image corresponds to a specific drug. The UwU-Net architecture was trained to predict 12 drug images from the 300-channel hyperspectral images. Figure 4.2 shows the results of these predictions and the corresponding 1 m/z bin image that contains the monoisotopic peak. While some of these low mass resolution bins are already highly correlated with the specific monoisotopic peak (e.g. Ipratropium and Vatalanib in Figures 4.2a and 4.2b, respectively), other images have strong background contributions and or conflicting drug spot signal due to fragment peaks from other drugs (e.g. Erlotinib and Gefitinib in Figures 4.2c and 4.2f, respectively). From Figure 4.2, it is apparent that the deep learning algorithm is able to reliably predict each drug's location from the low-resolution hyperspectral data even when there are conflicting background/fragment peaks or when the drug concentration is low (as in Lapatinib and Trametinib in Figures 4.2k and 4.2l). Even in Trametinib, where the drug is near the sensitivity limit for this MSI experiment, the UwU-Net correctly predicts the spot where the drug is present. Though the exact pixels predicted do not cleanly match (as noted by the PCC values for Trametinib in Table 4.2), the grouping of these sparse pixels in the correct spots suggest that the UwU-Net is picking-up the relevant spectral and spatial components for prediction.

Table 4.2. Quality Metric Values for the MSI Dataset Predictions

Drug (mass, m/z)	1 m/z bin		U-Nets (Non-HS)		UwU-Net (12-U)		UwU-Net (1-U, only drug bins)		UwU-Net (1-U)		UwU-Net (5-U)	
Ipratropium (332.223)	0.99	0.003	0.99	0.011	0.99	0.013	0.99	0.007	0.99	0.014	0.99	0.013
Vatalanib (347.107)	0.97	0.010	0.98	0.019	0.95	0.012	0.97	0.010	0.95	0.013	0.96	0.013

Erlotinib (394.177)	0.38	0.104	0.93	0.059	0.89	0.025	0.94	0.023	0.93	0.022	0.93	0.022
Sunitinib (399.220)	0.08	0.096	0.12	0.399	0.67	0.055	0.86	0.033	0.89	0.029	0.89	0.030
Pazopanib (438.171)	0.63	0.076	0.99	0.030	0.97	0.013	0.98	0.014	0.98	0.011	0.98	0.013
Gefitinib (447.160)	0.24	0.250	0.84	0.053	0.88	0.026	0.56	0.044	0.88	0.027	0.89	0.022
Sorafenib (465.094)	0.21	0.077	0.93	0.121	0.93	0.021	0.93	0.020	0.93	0.020	0.992	0.021
Dasatinib (488.267)	0.91	0.033	0.99	0.016	0.98	0.014	0.98	0.014	0.98	0.013	0.98	0.012
Imatinib (494.267)	0.25	0.240	0.72	0.096	0.59	0.057	0.73	0.042	0.78	0.040	0.75	0.039
Dabrafinib (520.143)	0.35	0.295	0.96	0.058	0.92	0.027	0.95	0.024	0.96	0.020	0.96	0.019
Lapatinib (581.143)	0.26	0.084	0.8	0.024	0.80	0.023	0.76	0.026	0.80	0.024	0.74	0.024
Trametinib (616.086)	0.05	0.047	0.11	0.023	0.12	0.030	0.03	0.024	0.10	0.025	0.24	0.023
PCC	0.44 ± 0.34		0.78 ± 0.32		0.81 ± 0.25		0.81 ± 0.28		0.85 ± 0.25		**0.85 ± 0.21**	
NRMSE	0.110 ± 0.098		0.076 ± 0.11		0.026 ± 0.015		0.023 ± 0.012		0.021 ± 0.008		**0.021 ± 0.007**	

The table shows spiked drugs, their respective masses, and the Pearson correlation coefficients (PCC, left column under each model) and normalized root mean squared error (NRMSE, right column under each model) for the low resolution and predicted images from various models with respect to the high resolution image for the drug. U-Nets (Non-HS) refers to individual traditional U-Nets trained from a single image input of low mass resolution (i.e. non-hyperspectral images). The "only drug bins" UwU-Net was trained on a 12 image input stack of only the relevant 1m/z images that contain the drug peak. All other UwU-Nets were trained using the full 300 image stack with various numbers of spatial U-Nets at their center (1-U, 5-U, or 12-U). The uncertainty values refer to the standard deviation among the respective metrics for the given model (n=12 for all).

To better understand the role of spectral and spatial learning in the UwU-Net, other U-Net and UwU-Net models were trained on this data with some varying parameters and compared in Table 4.2. To first understand the role of spectral vs spatial learning and their interplay on model accuracy, multiple basic U-Nets were trained on a single drug at a time. Here the single 1 m/z bin image and corresponding high mass resolution peak image were used for training. While some of the drugs are correctly identified and predicted (suggesting spatial learning of a single image from the hyperspectral stack may drive some drugs' predictions), many of the drugs (sunitinib, gefitinib, sorafenib, dabrafinib, and trametinib) go partially or entirely unpredicted. A single basic U-Net modified to accept 300 channels and output 12 channels again produces unacceptable results (Appendix C Figure A.C.1). The use of a UwU-Net with a single spatial U-Net at its center (denoted as 1-U in Table 4.2) allows for spectral learning of the data in addition to spatial learning.

When a stack of just the 12 drug 1 m/z bins is used for training (1-U, only drug bins in Table 4.2), only gefitinib, dabrafinib, and trametinib were unidentified. The use of the full 300 hyperspectral stack in the 1-U UwU-Net shows further improvement leaving only one spot of dabrafenib unpredicted. This suggests additional spectral information improves the accuracy of the model in drugs where spatial information from the principal bins is insufficient for prediction. The use of a 12-U UwU-Net on the full hyperspectral data eliminates any unidentified drug spots, but errantly predicts spots in sunitinib and initinib that do not exist in the respective truth images. A 5-U UwU-Net demonstrates the most accurate prediction of drug spots with no missing or errantly predicted spots for any of the 12 drugs (as seen in Figure 4.2). This analysis and comparison suggest that, like "depth" in a traditional U-Net or ResNet, architecture parameters such as spectral depth or number of spatial U-Nets at center can be empirically tuned to improve model accuracy.

These results highlight a capability of the UwU-Net to mine MSI datasets for relevant features from both spatial and rich spectral features afforded in MSI in a convolutional manner. One way this is potentially useful for MSI is in the design and execution of experiments. If *a priori* ground-truth information is available (in this case, the masses of the drug molecules sought, their locations, and their concentrations), a UwU-Net model can be trained and utilized in other similar experiments to vastly improve analysis speed. For example, while the training of this algorithm took ~8 hours, the final prediction of all images shown takes only ~1 second. This upfront single-time investment of training then affords analysis of further samples to be performed extremely quickly in comparison to costly linear analysis of each dataset. The specific demonstration presented here could also be highly useful for the miniaturization of MSI systems for *in situ* use where the tradeoff of reduced mass resolution would be mitigated by a pretrained algorithm. We also note the possibility of combining MSI with an orthogonal method such as fluorescence or

Raman imaging, to predict alternate imaging modalities using the UwU-Net as we demonstrate below.

4.3.3 *Label-free Organelle Fluorescence Prediction from SRS Microscopy Images*

Label-free prediction via deep learning has been a recent area of interest for augmenting the information acquired from a given microscopy modality[155]. The label-free prediction usually involves a microscopy image, such as transmitted light or autofluorescence microscopy, being converted to an image that mimics a more complex label-requisite modality like fluorescent or histologically stained images[120,129,156]. The value of this type of work is clear due to the elimination of staining protocols and the disadvantages associated with labeling the sample (photobleaching, toxicity, disruption of biological structures or functions, etc.). However, the quality of label-free prediction depends heavily on the information present in the input images[157]. For example, while transmitted-light microscopy is relatively simple to perform, it only reveals information based on light scattering due to differences in refractive index. In the context of cells and their organelles, there may not be significant enough difference between an organelle and cytosol to produce relevant information for a deep learning algorithm to reliably predict a corresponding organelle's fluorescence.

Compared to simple bright field or autofluorescence imaging, Raman imaging is a much more information-rich, label-free alternative. The Raman spectrum of a sample reflects specific molecular vibrations quantitatively associated with the molecules within. Hyperspectral SRS imaging improves the conventional Raman imaging by significantly speeding up the image acquisition by 3-4 orders of magnitude[34,35,38]. Regardless of the acquisition method, for biological samples, the Raman spectra are often congested and highly convolved due to the overlapping Raman bands from many different molecules. Principle component analysis and

phasor analysis have been used to extract individual organelles from the myriad of vibrational signatures in a cell[15,120]. However, the subtle variations of Raman spectra for individual organelles present significant challenges to the analysis of smaller structures such as mitochondria and endoplasmic reticulum (ER). Previous attempts to produce label-free staining based on hyperspectral Raman imaging have shown promising results for some organelles but not as rich of predictions for smaller ones[120]. The architecture we present here shows improved fluorescence prediction across 3 organelles. Deep learning using the rich spectral and spatial information afforded by hyperspectral SRS microscopy also outperforms previous work of label-free prediction from transmitted light microscopy[129].

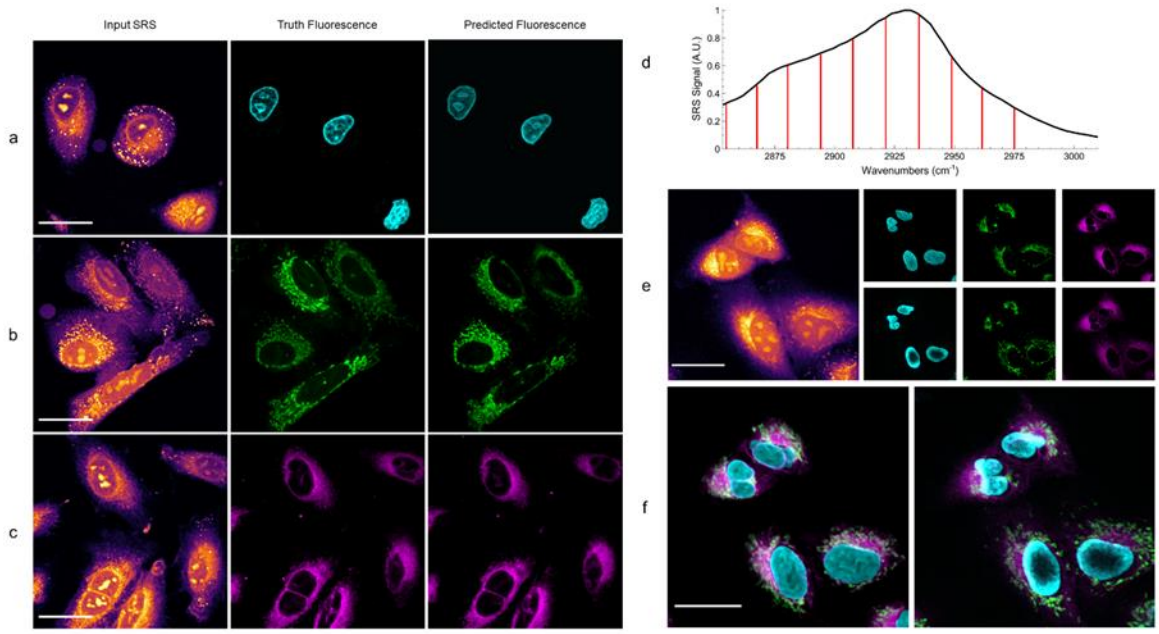

Figure 4.3. Predicted organelle fluorescence from hyperspectral SRS microscopy images All SRS images shown depict only the peak signal image from the hyperspectral stack. Panel a shows the prediction of nucleus fluorescence. Panel b shows the prediction of mitochondria fluorescence. Panel c shows the prediction of endoplasmic reticulum fluorescence. Panel d shows a typical cellular SRS spectrum (black) and the 10 vibrational transitions imaged and used for prediction (red). Note that the transitions marked in red represent the center of a band of probed transitions with a resolution of 19 cm^{-1}. The 15 cm^{-1} steps between each spectral image means the

entire CH vibrational region is effectively probed during hyperspectral imaging. Panel e shows an SRS image of live cells (left) that contain no dye, each algorithms predicted fluorescence (right, top row), and fluorescence images taken after the cells are stained (right, bottom row). Panel f shows an overlaid combination of each organelle prediction (left), and the same group of cells after staining (right).

As shown in Figures 4.3a – 4.3c, we create label-free prediction algorithms for nuclei, mitochondria, and endoplasmic reticulum fluorescence in fixed lung cancer cells (A549, from ATCC).. The accuracy of the predictions is quantified in Table 4.3 by Pearson's correlation coefficient (PCC), normalized root mean squared error (NRMSE), and feature similarity index (FSIM)[150,158].

Table 4.3. Quality Metric Values for the Label-free Prediction of Organelle Fluorescence

Organelle Model	PCC	NRMSE	FSIM
Nucleus	0.92 ± 0.03	0.047 ± 0.022	0.89 ± 0.04
Mitochondria	0.84 ± 0.05	0.059 ± 0.019	0.93 ± 0.02
Endoplasmic Reticulum	0.94 ± 0.02	0.038 ± 0.016	0.92 ± 0.03

The table shows pearson correlation coefficients (PCC), normalized root mean squared error (NRMSE), and feature similarity index (FSIM) values for the 3 organelles predicted from hyperspectral SRS images. Numbers shown are based on the average of all withheld test images (9, 9, and 7 images for nucleus, mitochondria, and ER, respectively) of 512 x 512 pixels. Uncertainty refers to the standard deviation among the withheld test images.

Across all computed quality metrics, we find high correlation and acceptably low error between predicted images and their respective truths. Previous work reported PCC values of 0.58, 0.69, and 0.70 for DNA (nucleus), mitochondria, and endoplasmic reticulum, respectively[129]. Thus, we see a significant improvement in label-free organelle prediction with the information-rich hyperspectral SRS microscopy in comparison to bright field microscopy. A basic U-Net was again trained for comparison as seen in Appendix C Figure A.C.1. While this task was more

successful than in the previous demonstrations, unacceptable residual SRS features were also present in the image. For additional comparison to another modern architecture used for image reconstructions, a U-Net utilizing ResNet Blocks[131,159] was also trained to predict the organelles (Appendix C Figure A.C.2 and Table A.C.2). While the Res-U-Net showed slightly improved organelle predictions in comparison to previously reported results, the UwU-Net predictions still outperformed across all organelles and metrics.

The utilization of both spectral and spatial information is paramount towards demonstrating utility of this architecture. This is most clearly demonstrated in the mitochondria prediction model by the differentiation of the organelle from lipid droplets in the cell. In SRS images, lipids droplets appear as bright "dots" typically ~1 μm in size. This means they have a similar size and shape to mitochondria, yet the trained models have clearly learned to exclude such similar features. This suggests that the model is not simply searching for the spatial features in the image to isolate and predict, but likely utilizing both spatial and spectral information to determine the position of the desired organelles. To confirm this, a simple 2D U-Net was trained using the single brightest SRS image to predict the fluorescence image (Appendix C Figure A.C.3). While the PCC values demonstrated by this traditional U-Net training still outperform previous work (likely due to the higher input image quality with respect to transmitted-light microscopy), they slightly underperform the UwU-Net where spectral information augments the prediction capability (Appendix C Table A.C.3). Moreover, the 2D U-Net models predict some spurious features such as nucleoli (Appendix C Figure A.C.3) or lipid droplets (Appendix C Figure A.C.4) as they are incapable of seeing the difference in vibrational spectral information for such features.

Finally, to demonstrate the multiplexing capability of the trained algorithms, hyperspectral SRS images of live A549 cells with none of the dyes present are used to predict organelle

fluorescence in Figures 4.3e and 4.3f. Here new prediction models have been trained for live cells in a similar manner as in the fixed cells (Appendix C Figure A.C.5). However, instead of predicting based on SRS images of cells where the dye is present (such as in Figures 4.3a – 4.3c and Figure A.C.5), the live cells are first imaged with SRS when no dye is present (Figure 4.3e, left). The cells are then stained while still mounted on the microscope and reimaged with two-photon fluorescence to acquire reference fluorescence images (Figure 4.3e, right, bottom row). The stain-free SRS images are used to predict fluorescence images using the pretrained models (Figure 4.3e, right, top row) and overlaid for comparison against the reference images (Figure 4.3f). As shown in Figure 4.3f, the label-free prediction in live cells matches well with the truth fluorescence images. We do, however, note slight mismatches in fields of view and cellular shape. This is due to both the sample moving and focus changing slightly during the staining process while mounted on the microscope. Additionally, organelle movement and cellular reorganization between SRS and fluorescence imaging (~10 minutes) leads to mismatch of exact spatial features. Regardless of these differences, the UwU-Net demonstrates a firm ability in predicting label-free fluorescence of organelles from SRS images of live cells.

4.4 CONCLUSIONS

In this work we have presented UwU-Net, a new architecture for deep learning using hyperspectral images. The architecture is highly flexible in both the types of tasks it can perform (e.g. classification, segmentation, label-free prediction) and the types of hyperspectral images with which it is compatible (e.g. remote sensing, MSI, and SRS microscopy). Specifically, we show excellent performance of Indian Pines classification with 99.48% overall accuracy for all classifications. We also demonstrate successful drug location prediction in fixed tissue from MSI data from windowed and binned images. This highlights the capability to mine spectrally dense

MSI datasets using both spectral and spatial information and offers new possibilities for deep learning in MSI. Finally, we show improved label-free prediction of organelle fluorescence by using hyperspectral SRS microscopy. We note a significant improvement in nuclear, mitochondrial, and ER prediction correlation with respect to previous work by the use of the UwU-Net to interpret spectral and spatial information.

We further note that while all models were trained using randomized starting parameters and stochastic gradient descent to minimize mean squared error (MSE) between output and truth images, the architecture is easily amenable to transfer learning methods and more complex error functions for particular tasks. We also note that the UwU-Net architecture can potentially be used in a generative adversarial network (GAN) framework to perform an even broader class of tasks[26,160]. However, GAN training of a UwU-Net is not feasible currently given memory constraints.

Finally, while only a subset of tasks and imaging techniques are demonstrated here, we expect the UwU-Net to be broadly applicable or adaptable to any reasonably designed computer vision task involving a hyperspectral imaging technique with potential use in medical imaging, microscopy, and remote sensing.

Chapter 5. LABEL-FREE CLASSIFICATION OF IMMUNE CELLS FROM HYPERSPECTRAL STIMULATED RAMAN SCATTERING MICROSCOPY IMAGES

5.1 INTRODUCTION

T cells are an integral part of the adaptive immune system. The various types of T cells (e.g. cytotoxic, helper, regulatory) and their respective functions modulate dynamically through the life of a given cell and are an important point of study in immune response, vaccine development[161], and cancer treatment[162]. The differentiation of T cells from their progenitors is thought to be a combination of transcription factor activation, other signaling molecule activations, and stochastic development[163,164]. Irrespective of the exact mechanisms behind cell differentiation, once developed and differentiated, a given T cell's identity and fate is irreversible in the body (unless genetic or epigenetic reprogramming is performed). Identification and study of T cells rely on surface protein markers that are associated with T cell types and differentiations. Current methods for studying T cells typically rely on extraction, fluorescent antibody staining tagged to the identifying surface proteins, and separation via flow cytometry[165]. While these methods can readily identify and collect the varying T cell differentiations, they are not typically applicable to in situ study of T cell function and response. Microscopy based studies of T cells have recently shown promising results in understanding T cell selection and fate in situ[166,167], but they often involve fluorescent labeling may limit the observation breadth and duration. Specifically, breadth of targets is limited to a few fluorescent labels at a time by the color crosstalk between dye molecule while the duration is limited by the photobleaching of the dyes. In this work we demonstrate stimulated Raman scattering (SRS)

microscopy as a potential label-free alternative to study of T cells with capability towards in situ imaging.

In contrast to fluorescent labeling or brightfield scattering microscopy, SRS microscopy targets the endogenous vibrational chemical differences of a sample at high spatial resolution without the need fluorescent labeling. This offers a significant label-free tool towards advancing the study of immune cells in situ without concerns of modulating biological potentially associated with fluorescent staining or the above-mentioned challenges with fluorescence microscopy.

Unlike fluorescence microscopy, however, SRS microscopy is limited in its specificity with respect to identifying both specific biological structures and isolating vibrational transitions from specific molecules in particular frequency regions. The latter is particularly relevant when imaging in the so-called carbon-hydrogen (CH) stretching region (~2800 – 3000 cm^{-1}). However, CH region imaging is often desirable due to the abundance of CH molecules, and thus, signal strength in this vibrational region which produces higher quality (i.e., higher signal to noise ratio [SNR]) images. While not as chemically specific as the "fingerprint region" of the vibrational spectrum (~400-1700 cm^{-1}), the CH region still offers chemical information related mostly to the lipid and protein profiles of biological samples. At the single cell level, the ratio of protein to lipid signal, as SRS signal is inherently quantitative, is correlated to the metabolic phenotype irrespective of molecular specificity. To this end, deep learning methods have been utilized to offset this tradeoff between the CH region's high SNR and lower chemical specificity. For example, recent publications have shown the ability to create semantic segmentation maps of organelles within cells[127], classify cancer[116], and predict organelle fluorescence images[110] all from label-free SRS images in the CH region. With deep learning, the strengths of SRS imaging in the CH region are augmented to enhance analysis and the capabilities of SRS microscopy. Here,

we demonstrate how deep learning can be used to classify label-free T cells from hyperspectral SRS images. We compare our deep learning method with traditional machine learning classification based solely on Raman spectral information. This comparison demonstrates how deep learning outperforms traditional machine learning; likely by incorporating relevant spatial information and its nonlinear interplay with the spectral information. We also demonstrate the ability to predict T cell fluorescence in genetically labelled lymph node tissue and suggest a scheme for training in situ label-free predictions using a model based on antibody fluorescent stains.

5.2 METHODS

5.2.1 *T Cell Extraction and Separation*

Male and female mice, 8-12 weeks of age, were used in compliance with University of Washington Institutional Animal Care and Use Committee Guidelines. Spleens were harvested and a single cell suspension was generated by compressing the spleen between rough glass slides. This cell suspension was filtered through 40 μm mesh into FACS buffer (HBSS, 0.5% BSA, 10 mM HEPES, pH 7.4). Cells were centrifuged at 350g for 4 minutes, resuspended in FC Block (33% 2.4G2 supernatant in FACS buffer) and incubated on ice for 15-30 minutes. Cells were stained with antibodies at concentrations listed in the antibody table (See Appendix E Table A.E.1) for 15-20 minutes in FC Block. Cells were resuspended in FACS buffer and sorted with a BD FACS ARIA III (BD Biosciences). Cells were sorted as follows: Naïve CD8: CD8+CD62L+CD44-, Central Memory CD8: CD8+CD62L+CD44+, Effector Memory CD8: CD8+CD62L-CD44+, Naïve CD4: CD4+CD62L+CD44-, Central Memory CD4: CD4+CD62L+CD44+, Effector Memory CD4: CD4+CD62L-CD44+, Sorted cells were washed and transferred to poly-l-lysine coated 8 well chamber slides (ThermoFisher) and were centrifuged

at 300g for 3 minutes. Cells were fixed for 15 minutes at room temperature with 4% Paraformaldehyde (Electron Microscopy Sciences) in PBS, rinsed with PBS, and cover slipped.

5.2.2 *Preparation of Lymph Node Slices for Imaging*

Male and female mice, 8-12 weeks of age, were used in compliance with University of Washington Institutional Animal Care and Use Committee Guidelines. Axial and inguinal lymph nodes were dissected from mice as shown previously[168]. The lymph nodes were washed with pbs before being fixed in 4% paraformaldehyde for 1 hour. Any excess fat tissue around the lymph node was then removed. The lymph nodes were then embedded in agarose, being careful to orient the lymph nodes on their side so as to expose the B and T cell follicle zones. The embedded lymph nodes were then sliced using a vibratome set to 300 μm thick slices. Only slices from the center sections of the lymph node were used for imaging. For the YFP genetically encoded line, the lymph nodes were imaged without additional processing. For the antibody stained lymph nodes, the slices were placed in FC Block for 2 hours before CD3+:PE (T cell, 540/580 nm ex/em) and CD19+:eFluo450 (B cell, 405/450 nm ex/em) were added. The cells were allowed to stain for 16 hours. The slices were then placed in PBS for 90 minutes to wash any additional stain or dye aggregations. Finally, the slices were placed on a glass slide and cover slipped for imaging.

5.2.3 *SRS and Two-photon Fluorescence Imaging*

SRS and Two-photon imaging were performed as described in the above section (See Figure 2.1, and Chapter 4 section 4.2.2). Briefly, SRS microscopy was performed in the CH region at powers of 20 mW for both beams for the flow separated cells and genetically encoded lymph node slices. Powers were set to 10 mW for each beam for the antibody fluorescence images. The flow separated cells were confirmed to exhibit no significant fluorescence signal on a PMT. For

the genetically encoded fluorescence imaging, a single PMT equipped with a 520/40 nm band pass filter was used to collect fluorescence images. For the antibody staining images, two PMTs were placed above the objective with fluorescent light separated first by a 700 nm long pass dichroic, then by a 500 nm long pass dichroic. A 570/40 nm band pass filter and 450/30 nm band pass filter were used to collect fluorescence from the T and B cells, respectively. SRS was acquired in spectral focusing hyperspectral stacks as described above in the spectral range of 2850-3000 cm^{-1}.

5.2.4 *Data Processing and Deep Learning*

Images of the flow separated T Cells were normalized at their 2920 cm^{-1} peak and had their background subtracted. In total 352 images were acquired across all T cell differentiations from two different mice. Truth "images" were created as 1x6 "vector" 8-bit images where each pixel corresponded to a given differentiation (Effector Memory CD4 [CD4E], Memory CD4 [CD4M], Naïve CD4 [CD4N], Effector Memory CD8 [CD8M], Memory CD8 [CD8M], Naïve CD8 [CD8N]). The pixel value was set to 255 for the given class and 0 for the other classes.

Hyperspectral SRS images from the lymph node slices were used without additional processing. The fluorescence images were averaged to a single image and used as ground truth in the training of models.

Deep learning classification for the flow separated cells was performed using a modified ResNet-50 architecture where the initial layer has spectral convolutions to handle the input 36-image hyperspectral stack as shown in Figure 5.1. 252 images randomly selected images were used for training while 100 images were withheld for testing.

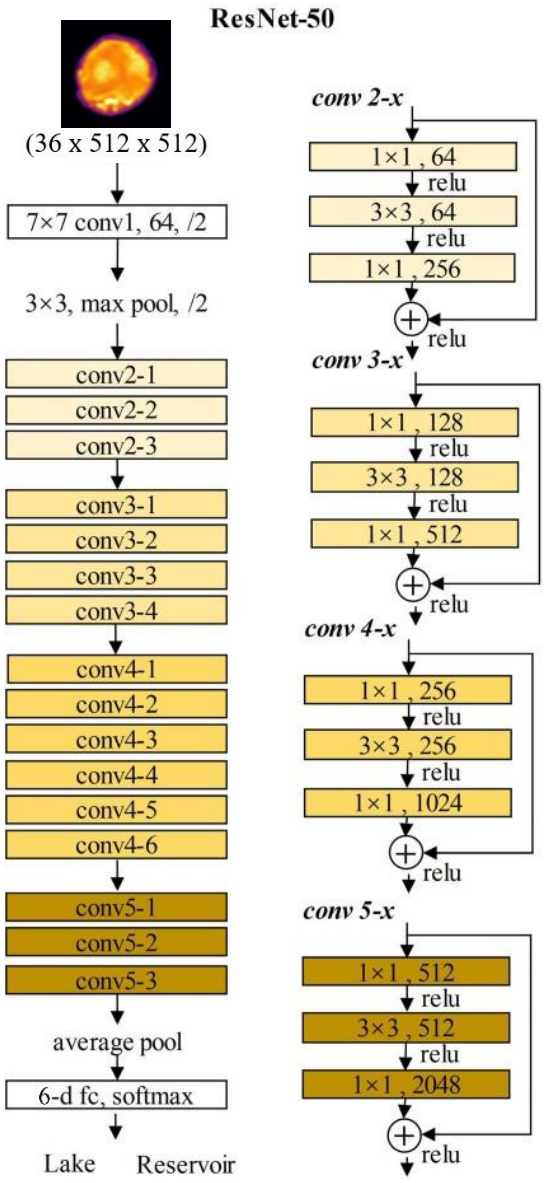

Figure 5.1. Architecture Diagrams and Indian Pines Classification

ResNet-50 architecture modified and used for deep learning classification of hyperspectral SRS T cell images. Adapted from reference [169] with permission.

For the lymph node slice image predictions a UwU-Net architecture (as shown in Figure 4.1) was used[110]. Here the default 80/20% test train split was used. The UwU-Net utilized 2 spectral steps with 6 central U-Nets, as described in Chapter 4.

All models were trained on a homebuilt Linux machine equipped with a Titan RTX GPU and AMD Threadripper 4 X399 CPU. The T cell classification and fluorescence prediction models took ~22 hours and ~14 hours to train respectively. Once trained, classification and prediction of the individual fields of view takes milliseconds.

5.3 Results and Discussion

5.3.1 *Hyperspectral SRS Images of Flow Cytometry Separated T Cells*

Representative SRS images of the different T cells and the average SRS spectra of the cells are shown in Figure 5.2. We note no obvious visual differences among the different types of T cells across the populations of ~40-50 cells in each class with the exception of CD4N cells which, in aggregate, exhibit less distinct sub cellular features (such as nuclei, nucleoli, lipid droplets, etc.).

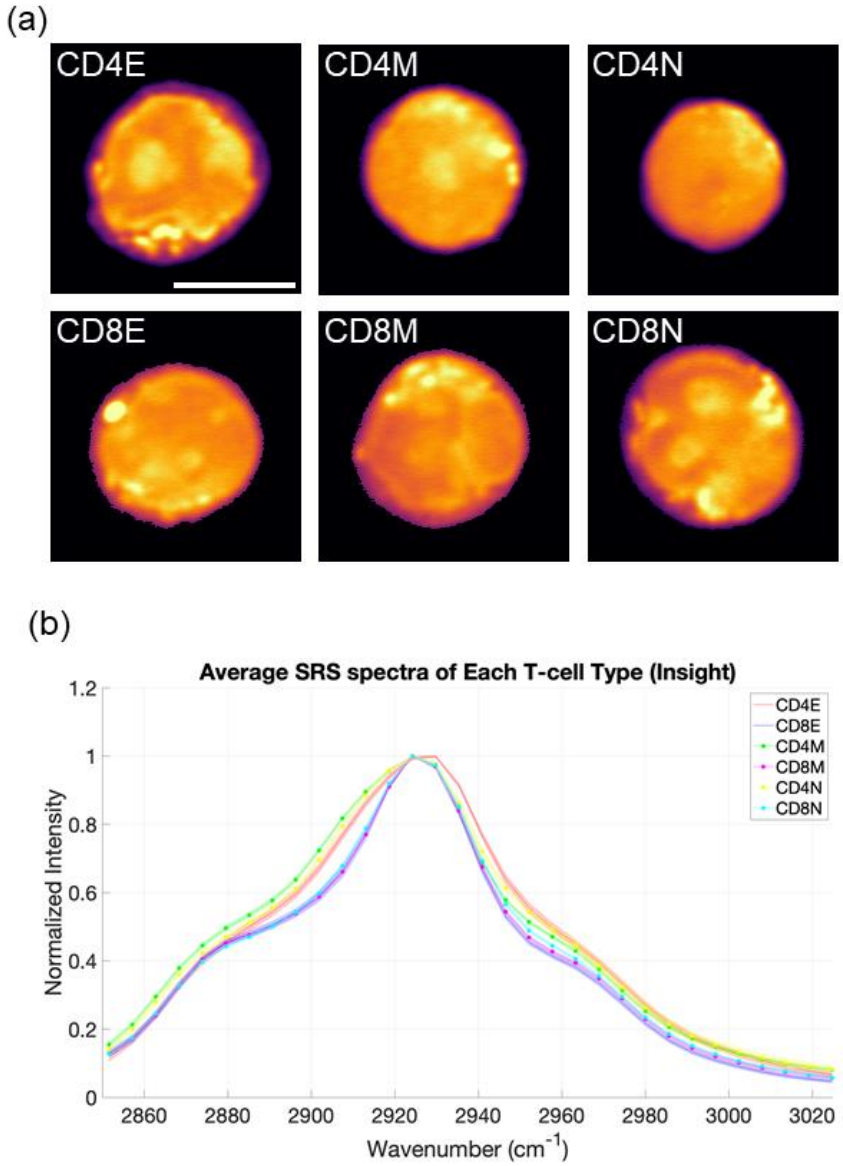

Figure 5.2. SRS Images and Spectra of T Cells

Panel a shows representative SRS images for each type of T Cell separated by flow. Where the image is an average projection of the entire SRS hyperspectral stack. Panel b shoes the average cellular SRS spectra across each differentiation normalized to the protein (2935 cm^{-1} peak). Standard deviations are drawn as splines into the spectra. Scalebar = 5 μm

While the overall spectral shape is consistent, there are subtle differences between the cell types especially with respect to CD4 vs CD8. These spectra indicate the chemical makeup of the

cell differentiations may provide leverage in creating a classification model based on the data. These chemical differences likely result from phenotypic differences in metabolism within the cells. Specifically, the ratio between the protein and lipid peaks within the cellular spectra may reflect more catabolic or anabolic states that T cells express depending on differentiation[170]. We have observed these spectral differences between CD4 and CD8 cells across the two different mice from which T cells were captured, as well as on a separate SRS microscope using broader band laser pulses.

5.3.2 *Machine Learning Classification of T Cells*

Here, the average cellular spectra are utilized to create a machine learning classifier. The normalized spectra of 252 training cells were explored with principle component analysis (PCA). The 1st and 2nd principle components capture ~97% of the variance in the data. If these principle components are mapped back to the spectral domain, we see that they are apparently related to the protein sidebands (2900 – 2950 cm^{-1}) and the lipid peak centered at 2850 cm^{-1} as seen in Figure 5.3a.

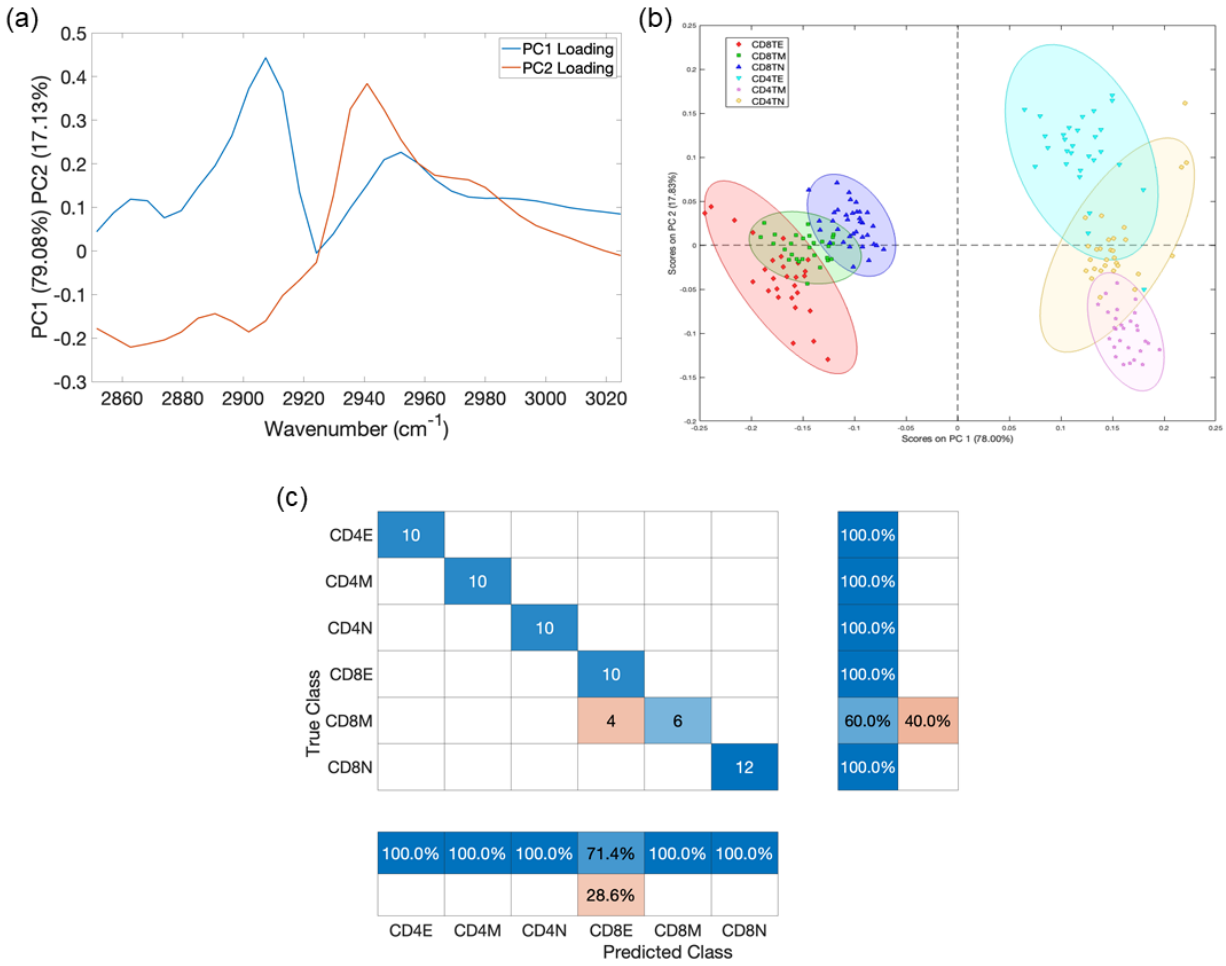

Figure 5.3. Machine Learning Results on T Cell Spectra

Panel a shows the loading of the first and second principle components from the T Cell data of all differentiations as mapped to the Raman spectrum. Panel b shows a simple clustering analysis of when the labeled cells are plotted with respect to their scores on the first and second principle components. Panel c shows the confusion matrix plots for the classifier based on PLS-DA of the withheld T Cell spectra. Overall Accuracy: 93.5%. CD4 vs CD8 Accuracy: 100%.

In addition to the loading plots of the principle components, the labeled T cell spectra can be mapped based on their scores within the principle components. Clustering analysis as shown in Figure 5.3B demonstrates that the CD4 and Cd8 groups of cells are highly separable based solely on their spectra, though there may be some overlap between the Effector/Memory/Naïve differentiations within each group. We then build a classifier based on this data using partial least

squares discriminant analysis (PLS-DA) and predict withheld test spectra as shown in Figure 5.3C. The classifier demonstrates an overall accuracy of 93.5% across the test dataset. This classifier is also able to, as expected, identify CD4 vs CD8 T cells based solely on their SRS spectra with 100% accuracy. The differentiations with CD4 cells are further also predicted with 100% accuracy. It is only the CD8-Effector-Memory and CD8-Memory T cells that exhibit confusion in their classification. Considering this model with reasonably high accuracy was trained only on, it stands to reason the spectral information as it relates to metabolic phenotype (here seen as the protein/lipid signal proportions in the spectra) may drive a large portion of a deep learning prediction which should perform at least as well.

5.3.3 *Deep Learning Classification of Hyperspectral SRS Images of T Cells*

Considering the reasonably accurate classification of T cells based on simple machine learning of their SRS spectra, we expect an equivalently rich deep learning classification model with further improved accuracy with the incorporation of spatial information. To this end, we used the same 252 T Cells in the training of our modified ResNet-50 (Figure 5.2). However, rather than only providing the 1-dimensional spectral information, we allowed the deep learning algorithm to convolve our 3-dimensional hyperspectral images. The trained model was then used to predict 100 withheld test images as depicted in Figure 5.4

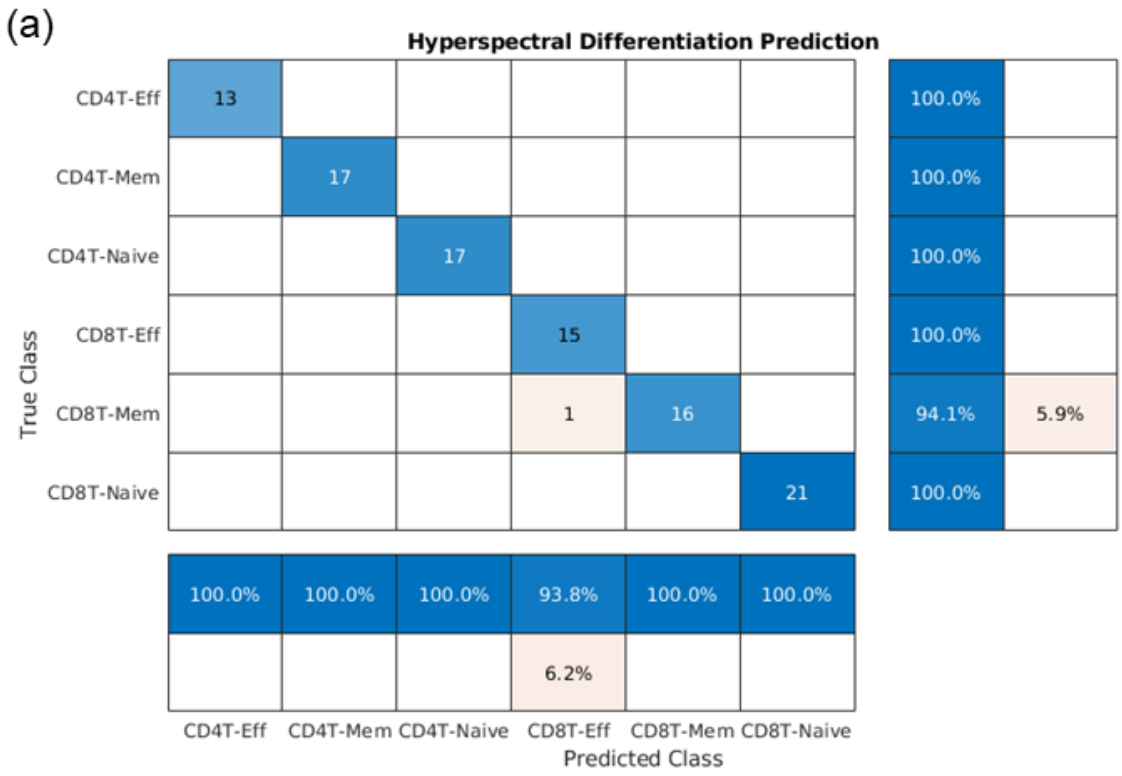

Figure 5.4. Deep Learning Classification of Hyperspectral SRS Images of T Cells

Panel a shows the confusion matrix plots for the classifier based on a modified ResNet-50 architecture for hyperspectral image classification into the 6 relevant classes. Overall Accuracy: 99% across the 100 test images. CD4 vs CD8 Accuracy: 100%.

We observe a continued 100% accuracy in classifying CD4 vs CD8 cells as was observed in the machine learning work. We also see an improvement in overall accuracy to 99% over all differentiations of T cells. We suggest this demonstrates the added benefits of both incorporating spatial information in the classifier and learning the nonlinear relationship between the input hyperspectral data and the output classes. Indeed, even though the cells are apparently visually indistinguishable, there may be subtle differences between the CD8-Eff and Cd8-Mem cells that allow for improved differentiation of the classes. In future work, the extent of this spectral vs spatial influence could be studied as was shown in the original UwU-Net paper (Chapter 4, [110])

with the mass spectrometry imaging data. That is, the number of free parameters associated with spectral and spatial features could be modulated to empirically determine optimum hyperparameters of the architecture. Further, the strongest demonstration and validation of our classifier would be imaging a mixture of T cells (all differentiations) in hyperspectral SRS microscopy, predicting their identity with our pretrained classifier, then staining and reimaging the same cells to confirm their truth identity.

5.3.4 *Deep Learning Prediction of Genetically Encoded Fluorescent T Cells in Lymph Node*

While classification of individual cell images suggests that SRS imaging can identify T cells based on label-free vibrational imaging, the value of performing this label-free classification lies in identifying cells in situ in tissue where long term, nondestructive imaging techniques could greatly benefit immunology. In this section we explore two possible methods towards building deep learning algorithms to identify and image specific immune cells within tissue. Specifically, we demonstrate the ability to predict T cell fluorescence in lymph node tissue from mice with YFP-encoded T cells. We also demonstrate the ability to acquire simultaneous SRS and two-photon fluorescence images of antibody-stained cells in wild type lymph node.

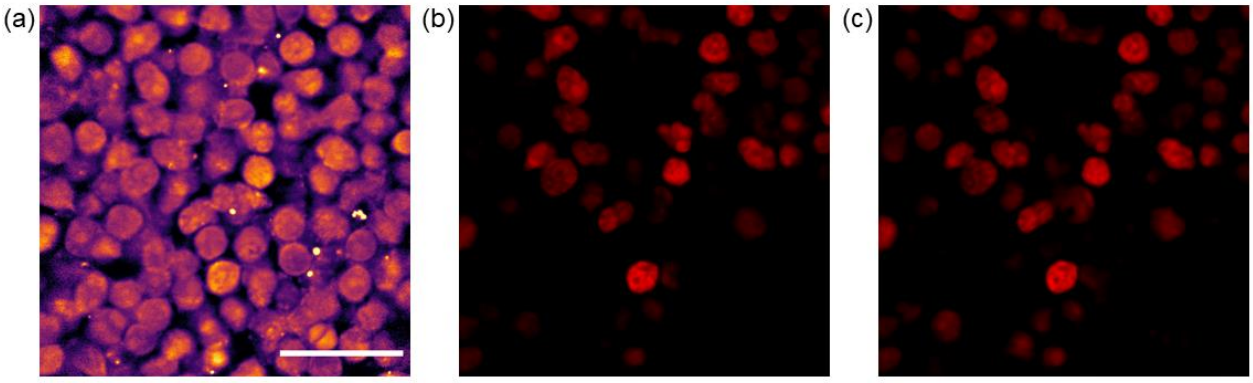

Figure 5.5. Deep Learning Prediction of T Cell Fluorescence from SRS Images

Panel a shows the average SRS image projection of the hyperspectral stack. The 36 spectral points interrogated are the same as those shown in Figure 6.2 for the T cell characterization. Panel b shows the same field of view in the fluorescence channel. Panel c shows the deep learning algorithm's prediction of T cell fluorescence from the input hyperspectral SRS image (panel a). Scalebar = 25 μm

Genetically encoded axial lymph nodes were first imaged under our simultaneous SRS and fluorescence microscope as described above. Here, 40 image pairs of SRS and fluorescence of various fields of view from the lymph node were used to train a UwU-Net model to predict fluorescence images from hyperspectral SRS images. As shown in the representative field of view in Figure 5.5a -5.5c, the model demonstrates a strong ability to identify and predict the fluorescent T cells in the SRS image. Across the test images (n = 10) a pearson correlation coefficient of 0.90 is achieved with respect to the truth fluorescence images. However, it is evident in the SRS images that some cells express so much YFP as they are noticeably bright even in the SRS image. This is likely due to some transient absorption signal leaking into the SRS image. This is a concerning feature for this particular method of imaging T cells in situ towards a label-free method. That is, the algorithm may be learning from the wrong signals (i.e. transient absorption of the encoded dye vs SRS signal). It is also not easily remedied, as it may be difficult to change the particular expression level of the YFP within these cells. We do not, however, that T cells with similar SRS signal to non-T cells also appear to be properly predicted. Further this fluorescence prediction method does not seem to have an obvious validation method to check our trained deep learning algorithm as in the UwU-Net organelle fluorescence prediction. Considering this significant hurdle in amending the fluorescence levels towards insensitivity in the SRS channel and in validating the output predictions of our model. We do not expect this to be a valuable target towards label-free in situ classification of immune cells.

Instead of the prediction of genetically encoded fluorescence to identify immune cells in situ. Antibody fluorescent staining may also provide a promising avenue towards a deep learning based label-free prediction and classification scheme. Indeed, for studies interested the study of the spatial distribution of immune cells antibody fluorescence has already been demonstrated[166,167,171,172]. It is theoretically possible to either directly predict this antibody stain from hyperspectral SRS images as was seen in the UwU-Net organelle fluorescence prediction, or to create ground truths for individually segmented SRS images of cells to be classified as shown above in the flow-separated cells. Within this scheme however, is a careful balance of staining a lymph node strongly enough to detect its fluorescence without influencing the SRS images as was observed in Figure 5.5. Our preliminary experiments to this end have demonstrated the difficulty in optimizing this staining process for our simultaneous SRS and two photon imaging though recent data suggest a strong possibility towards acquiring a training dataset for deep learning models as shown in Figure 5.6.

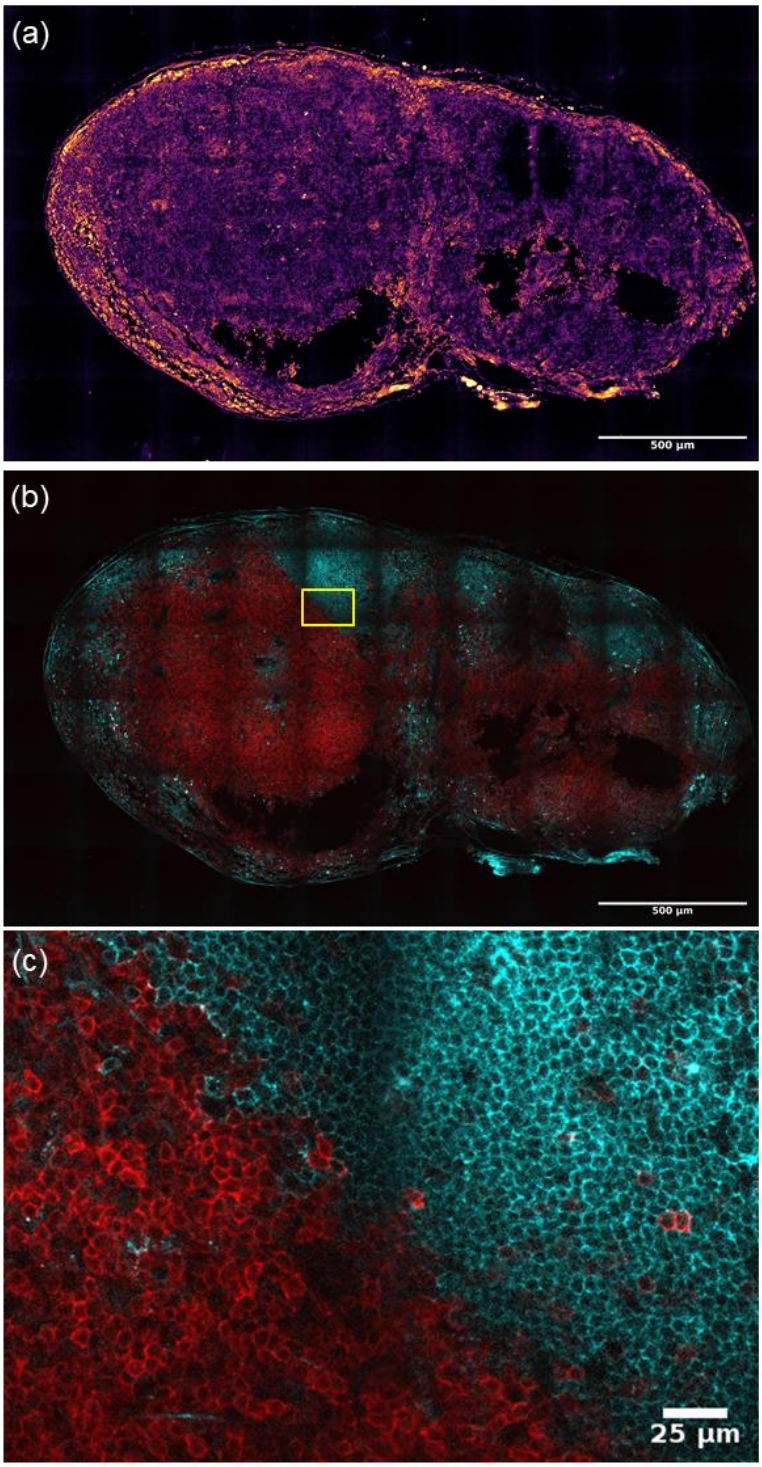

Figure 5.6. Simultaneous SRS imaging and Antibody Fluorescence Images of Lymph Node

Panel a shows the average SRS image projection of the hyperspectral stack across a fixed axial lymph node slice. Panel b shows the same field of view in both fluorescence channels. Here B cells are shown in cyan and T cells are shown in red. Scalebars for panels a and b are 500 µm.

Panel c shows a zoomed in area of panel b (highlighted in yellow in panel b) showing an area where the B cell follicle and T cell follicle meet. Scalebar = 25 μm

The SRS images of the the lymph node slice in Figure 5.6a show a relatively homogenous signal across the tissue with no observable fluorescence cross talk. Careful examination however reveals some field heterogeneity artifacts as is to be expected in such low power imaging (10 mW in both arms). The fluorescence channels in Figure 5.6b and the zoom in shown in Figure 5.6c do demonstrate the expected B cell (cyan) follicle and T cell (red) follicle zones in the expected places showing a promising method for identifying B vs T cells in lymph node tissue. However, because of the weak powers and stingy overlapping of tiles, field heterogeneity effects significantly affect the overall stitching of lymph node images. The zoom in area of Figure 5.6c shows the meeting area between the B and T cell follicles demonstrating the specificity of the staining and imaging with no fluorescence crosstalk. It also shows the lower SNR of these fields of view near adjacent image edges. Preliminary deep learning predictions based on this dataset are not particularly rich in their prediction, especially outside of the very center of images. Additionally, outside the center of SRS images, cells are difficult to distinguish from the background making segmentation of individual cells in this dataset difficult. Imminent experiments will repeat this promising staining protocol with higher powers and a higher magnification objective to address the current shortcomings in training a robust deep learning model for B and T cell prediction in lymph node images. Future work may also utilize the antibody staining rather as a guide for segmentation and classification of individual cells within the tissue, or to create semantic segmentation maps of T vs B cell zones in lymph nodes.

5.4 Conclusions and Future work

We have demonstrated how hyperspectral SRS images in combination with deep learning algorithms can be used to study T cells as a potentially label-free alternative to traditional fluorescence microscopy and flow-cytometry based experiments. Specifically, we have shown that deep learning can be used in conjunction with SRS imaging to produce high quality models that can predict T cell differentiation identity with 99% accuracy and fluorescence of T cells in lymph nodes with fidelity. This classification appears to be largely driven by spectral differences between the differentiations of T cells which relates to their metabolic phenotype. We also demonstrate two exciting methods to produce imaging datasets of immune cells in lymph nodes that can be used to train label-free prediction algorithms of immune cells from hyperspectral SRS imaging. While genetically encoded fluorescence may be improperly predicting to transient absorption signal, antibody fluorescent staining has recently shown promise as a valuable target towards the goal of label-free deep learning classification algorithm training. While this work demonstrates a sound foundation for studying immune cells in a label-free microscopy modality, future work will be needed to validate the pretrained classification models and to observe and classify cells in situ label-free.

Chapter 6. LABEL-FREE CLASSIFICATION OF THYROID CANCER FROM HYPERSPECTRAL STIMULATED RAMAN SCATTERING MICROSCOPY IMAGES

6.1 INTRODUCTION

Thyroid cancer is the most common endocrine system cancer an estimated 43,800 new cases diagnosed in 2022[173]. Often, an initial thyroid nodule is discovered during routine checkups. After some initial waiting period to determine a nodule's growth and potential malignancy, a nodule may be further evaluated by ultrasound-guided fine needle aspiration (FNA) biopsy. This biopsy uses a small gauge needle to directly stick and extract cells from the suspicious nodule without the need for invasive surgery. The FNA sample is then interpreted by specialized pathologists to more accurately asses and diagnose the nodule following the "Bethesda System for Reporting Thyroid Cytopathology"[174]. This system categorizes FNA biopsy samples as: 1. Nondiagnostic, 2. Benign, 3. Atypia of undetermined significance, 4. Follicular neoplasm, 5. Suspicious for malignancy, and 6. Malignant. However, cytopathology may not give a full picture of a thyroid nodule's status as 15-30% of FNA biopsy analysis are considered "indeterminate"[175]. This indeterminate form covers the latter 3 categories of the Bethesda system and as such are carefully assessed further for potential malignancy. Unfortunately, many of the features that distinguish such latter categories of thyroid nodules are based on spatial features of the nodule in the context of the whole tissue which can only be properly assess with invasive surgery, biopsy, and traditional pathology work up. Further, given the potential malignancy of such cases, thyroidectomy or partial thyroidectomy is often performed in these cases anyway, though it has been shown that as many as 70-80% of nodules excised and evaluated are ultimately

diagnosed as benign[176]. These potentially unnecessary thyroidectomies put a serious burden on patient quality of life as daily medication is required to balance the hormonal production of a healthy thyroid, and complications from surgery such as nerve damage, trouble swallowing, or trouble speaking are common.

As such recent works have offered potential alternatives to traditional cytopathology on FNA samples to improve their minimally invasive diagnostic accuracy and minimize unnecessary surgeries. For example, genetic sequencing assays have recently been developed and adopted to further asses and diagnose FNA sampled nodules. While these have been helpful in improving diagnostic accuracy and enhancing treatment planning, they're predictive power may still be lacking[177]. Other recent works have utilized machine learning and deep learning models on thyroid images to diagnose thyroid nodules. For example, Halicek et al. show deep learning on hyperspectral reflectance images of thyroid tissue with up to 79% overall accuracy[178]. Oliveira et al. show hyperspectral spontaneous Raman scattering microscopy as a promising imaging method to diagnose thyroid cancer in combination with machine learning achieving 97% accuracy in discriminating classic variant papillary thyroid carcinoma and benign nodules[179,180]. It is worth noting that the latter works are based solely on the Raman spectra of the imaged cells. While the apparent accuracy is high and related to putative chemical differences (e.g. phenylalanine's peak at 1000 cm^{-1} being much stronger in particular cancer types), throughput becomes an issue with imaging individual cells with spontaneous Raman microscopy. For example, even with line scanning, Oliveira et al report imaging times into the tens of minutes per single cell. Here we show SRS microscopy in combination with a deep learning classifier may be a high throughput alternative to the previously demonstrated spontaneous Raman microscopy.

6.2 METHODS

6.2.1 *Thyroid Sample Preparation*

Thyroid nodule samples were collected from thyroidectomy patients through NWBiosamples. In addition to the suspicious nodules, normal thyroid was also collected from the same patient if possible to be used as a control sample for deep learning training. The collected samples were first chopped finely with a razor blade before dissociating in collagenase (Worthington type II) overnight at 37°C. The dissociated tissue was then filtered through a 70 μm mesh filter and centrifuged at 250 G. The cell pellet was then resuspended in formalin solution and allowed to fix for 20 minutes. The cells were then centrifuged again and resuspended in fresh PBS. This wash step was repeated once more before imaging. Finally, the pellet of fixed cells was suspended in a small amount of PBS. A small amount of the cell suspension was sandwiched between a glass slide and coverslip and imaged.

6.2.2 *SRS Imaging*

SRS imaging was performed as described before in combination with a parabolic fiber amplifier[8]. SRS images of the thyroid samples were taken at 3 different spectral zones by tuning the pump laser color: 900 – 1200 cm^{-1}, 1350 – 1720 cm^{-1}, and 2850 – 3030 cm^{-1}. The two fingerprint regions were chosen based on the putative distinguishing Raman peaks demonstrated by Olveira et al. The CH region was additionally taken given the high SNR of the images in comparison to the fingerprint images.

6.2.3 *Deep Learning*

Deep learning classification for the dissociated thyroid cells was performed using a modified ResNet-50 architecture where the initial layer has spectral convolutions to handle the

input various hyperspectral stacks similar to the architecture shown in Figure 5.1. Only the input and output channels were changed to suit the thyroid datasets. That is, 54, 65, or 36 input channels for the three preciously listed spectral zones, respectively. And either 2 output classes (for normal vs cancer classification) or 7 output classes (for simultaneous classification of all potential cancer types.

With respect to computing hardware, models were trained on a homebuilt Linux machine equipped with a Titan RTX GPU and AMD Threadripper 4 X399 CPU. The single cell prediction models were trained within 3 hours. Whole field of view models were trained in 20 hours. Trained models were able to predict fields of view within milliseconds.

6.3 Results and Discussion

6.3.1 *Machine Learning Results on SRS Spectra from Thyroid*

Hyperspectral SRS images of both tumor and normal cells revealed some similar spectral features in the fingerprint region as reported by Oliveira et al. as shown in Figure 6.1a – 6.1d. However, given the lower sensitivity and thus lower SNR in the fingerprint region, attempts at machine learning classification based on these images were not particularly rich (accuracy ~60%). This low accuracy is also potentially due to slight spectral shifts in our lasers from one day to another over the course of multiple years during case collection. The fingerprint region is particularly sensitive to these spectral shifts, while the CH region has been shown to be far more stable and spectrally consistent over a long time. In addition to spectral stability, CH region offers a greater abundance of SRS signal thus affording higher SNR images. As such, we have chosen to focus classification on the hyperspectral SRS images taken in the CH region.

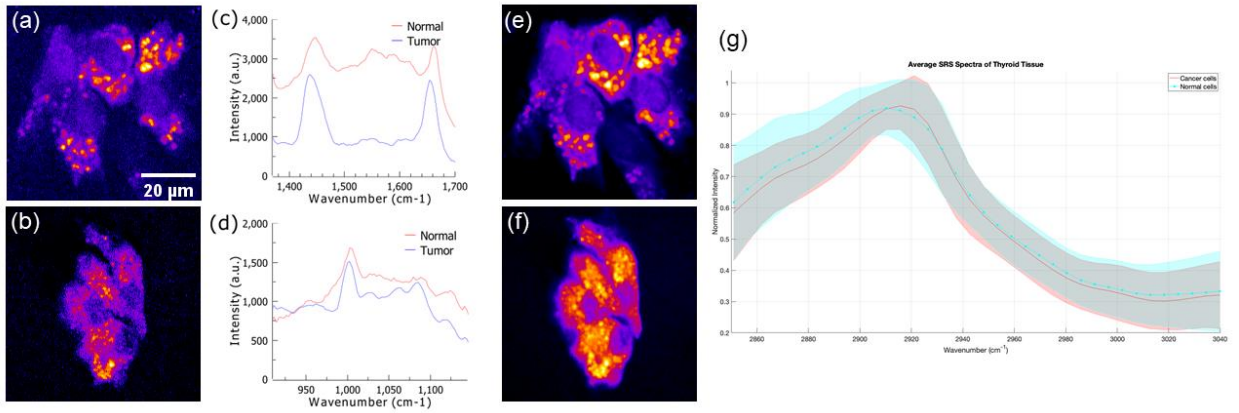

Figure 6.1. Hyperspectral SRS imaging of Dissociated Thyroid Tissue

Panel a normal thyroid cells imaged in the ~1000 cm^{-1} spectral regime. Panel b shows cancerous thyroid cells imaged in the same ~1000 cm^{-1} spectral regime. The fingerprint plots of normal vs tumor cells for this particular case are shown in panels c and d. Panel e and Panel f show the CH region SRS images of the same fields of view as in panels a and b. Panel G shows the average CH SRS spectrum across all cases for both the cancer samples (red) and normal samples (cyan).

The low sensitivity of the fingerprint regions is observed in Figure 6.1a and Figure 6.1b where images appear grainy and noisy exhibiting lower SNR. Further, given the lower signal levels, background becomes an important consideration especially in the SRS spectrum across the cells seen in Figures 6.1c and 6.1d. While the tumor exhibits slightly stronger peaks at 1000 cm^{-1} and 1650 cm^{-1} as expected (relative to the background), normal spectrum is clearly influenced by parasitic background signal likely from cross-phase modulation reducing the signal to background ratio significantly. This lower SBR can also be seen in Figure 6.1a with the off cell areas also exhibiting some signal. In contrast to these noisier images, the CH region images of the same fields of view shown in Figures 6.1e and 6.1f exhibit distinct signal from the background and higher SBR. Though previous work only examined the fingerprint regime, it is plausible that chemical differences between normal and tumor cells would also be exhibited in the CH region due to

differences in lipid and protein content within cells. Cancer cells might be expected to on average contain fewer lipids and more proteins given the increased metabolic rate and anomalously high growth rate. A slight trend in this direction is observed in Figure 6.1g, though across the preponderance of cases statistical significance is not well established. Despite this there may still be nonlinear trends between both the spectral and spatial information that could provide insight to a classification model when deep learning is utilized.

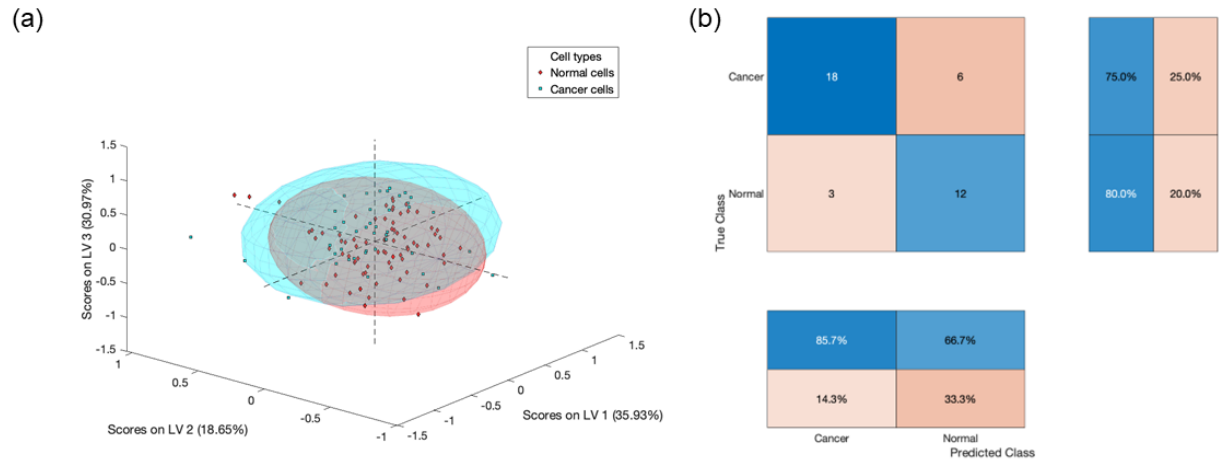

Figure 6.2. Machine Learning Classification based on SRS CH Spectra

Panel a shows the population of individual FOV's plotted by the principle components assigned from the spectra of each FOV. PLS-LDA is used to group the populations based on their labeling as either normal or cancer cells. Panel b shows the confusion matrix for the classifier on the test dataset with an overall accuracy of 76.9%, sensitivity of 75.0%, and specificity of 80.0%.

We first perform PCA on the spectra from the thyroid cases and build a discrimination model based on PLS-DA as shown in Chapter 5. As shown in Figure 6.2a, the cancer and noncancer cells are not particularly well separated in the principle component space, however a classifier is still able to assign labels with an overall accuracy of 76.9% as seen in Figure 6.2b. While better than the models based on the fingerprint spectra, it is still far below the expected

accuracy with respect to previous work utilizing similar methods. It is possible here that the high error is due to both spectral inconsistency amongst cases as well as noisy labeling. The latter is an interesting concern given that a dissociated sample is taken to be either completely normal or completely cancer across all its images which may contain hundreds of individual cells. It seems mostly improbable that a tumor section would contain no healthy cells after a conservative resection during the thyroidectomy. The normal tissue, however, is often taken far from the known nodule of interest, and is more likely to contain only normal cells perhaps explaining the slightly higher specificity here. Mislabeling of normal tissue is still a possibility, though, depending on the progression of the thyroid cancer state.

6.3.2 *Deep Learning Results on Hyperspectral SRS Images of Thyroid*

In contrast to machine learning, we also trained deep learning models that were able to utilize the spatial information of the hyperspectral images in addition to their chemical spectra. Training a deep learning model across whole fields of view for all cases, reasonably high accuracies are achieved for both classifying cancer vs not cancer and classifying the possible types of cancer. Figure 6.2a shows an overall accuracy of 96.4% benign vs cancer prediction with relatively balanced false positive (n = 3) and false negative (n = 2) for all available fields of view. An accuracy of 92.1% is achieved for prediction across all classes with the most confusion coming from over predicting the PTC class. This is expected given the relative imbalance of training data strongly favoring PTC. In Figure 6.2b, the diagnostic prediction of cases is created by summing the total predictive values from the individual fields of view. Here an accuracy of 100% is achieved for benign vs cancer cases and an overall accuracy of 95.5% is observed across all possible classifications. One important caveat here, however, is that these results include a mixture of

testing and training data given the sparse number of cases of the rarer thyroid cancer types. This consideration notwithstanding, we do observe similar error levels between the testing data (93% accuracy by field of view) and training data (95% accuracy by field of view) with respect to benign vs cancer prediction. We thus do not suspect the model of overfitting to the training data. We expect for these data to become more statistically robust as we continue to acquire more thyroid cancer cases aside from PTC diagnoses.

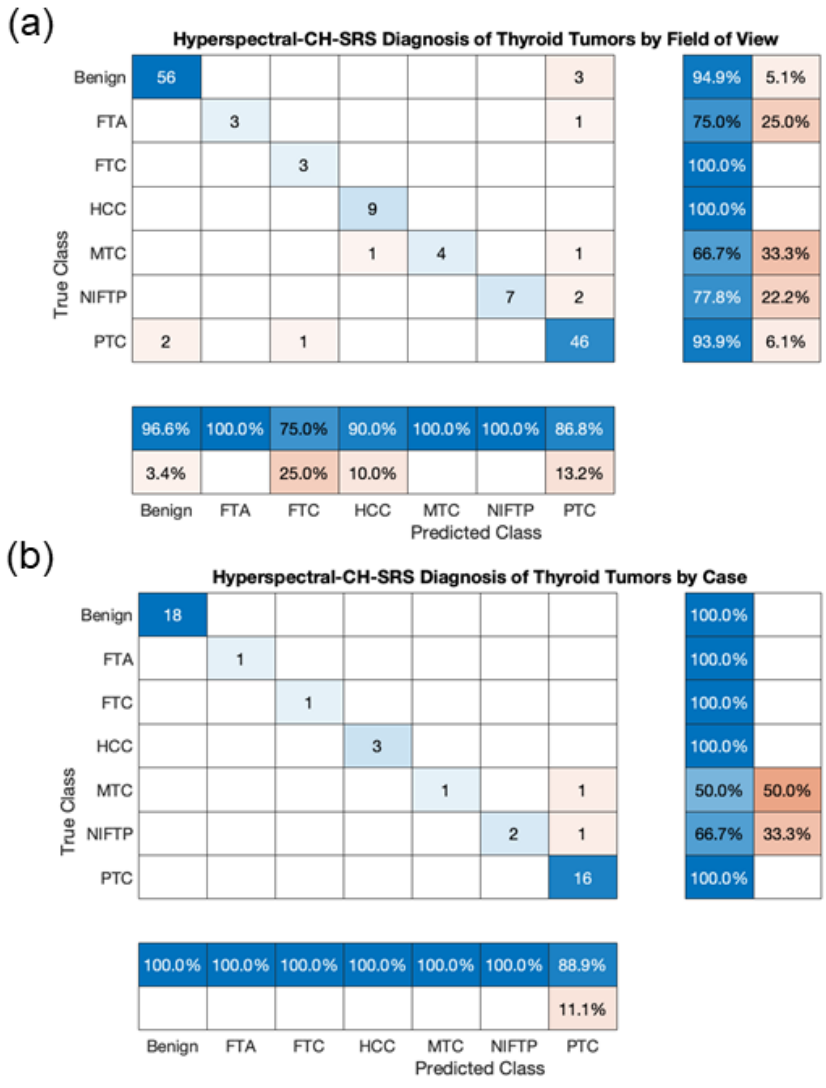

Figure 6.3. Hyperspectral SRS imaging of Dissociated Thyroid Tissue

Panel a shows the confusion matrix of ResNet-50 predicted classifications of thyroid samples by field of view across all 7 possible classes. Benign vs Cancer accuracy: 96.4%. Overall Accuracy: 92.1%. Panel b shows the confusion matrix of ResNet-50 predicted classifications by patient case. This is achieved by summing the predictive values across each cases fields of view. Benign vs Cancer accuracy: 100%. Overall Accuracy: 95.5%. These confusion matrixes include predictions from both the training and testing datasets given the still limited number of total cases and images. FTA: Follicular Thyroid Adenoma, FTC: Follicular Thyroid Carcinoma, HCC: Hurthel Cell Carcinoma, MTC: Medullary Thyroid Carcinoma, NIFTP, Noninvasive Follicular Thyroid with Paippilary features. PTC: Papillary Thyroid Carcinoma.

6.4 Conclusions and Future Work

Here we showed a hyperspectral SRS microscopy-based method for analyzing FNA biopsy like samples. In combination with the training of a deep learning algorithm that interprets both the spectral and spatial information within the SRS images, promising diagnostic accuracies are achieved. However, the current data is slightly imbalanced towards the most common type of thyroid cancer, PTC. Future work will see the continued collection of cases with a focus on more cases from the other types of thyroid cancer here explored. Another active area of work is training and predicting based on individual cells from the collected fields of view which can often contain 20-50 cells each. This involves segmentation of cells from all the fields of view then slightly modified training parameters to build a more statistically robust model. This is a promising end as it highlights the higher throughput potential of SRS microscopy in comparison to the previously demonstrated spontaneous Raman microscopy-based predictions. In all we expect this work to show a valuable tool towards translational diagnosis of thyroid cancer from minimally invasive FNA samples.

Chapter 7. CONCLUSIONS AND FUTURE DIRECTIONS

In all, this work has strived to present SRS microscopy as a valuable tool for biophotonics when used in combination with modern deep learning techniques. Specifically, we have shown how deep learning and augment the technical shortcomings of SRS microscopy. For example, in our denoising work[27] we show significant restoration of signal features from near limit levels which may unlock avenues of lower power SRS imaging, or advanced imaging arrangements such as a potential light-sheet SRS modality (where signal might expected to be lower due to the looser focusing of the pump and Stokes beams). We also show that our deep learning algorithms have practical value in extending the imaging depth limit of SRS microscopy where signal deep into tissue is limited by scatter, lowering the effective power at focus. These technical advancements through deep learning can also facilitate faster scanning methods where the inherent trade off is signal to noise ratio as shorter and shorter effective averaging of lock-in signal at each pixel is necessary. Though recovery of spatial features is shown here, we also expect a similar arrangement is feasible for recovery of spectral features given recent publications towards this end [127,181]. Though technically difficult to achieve, multiple coaligned beam lines would be necessary, one could imagine training femtosecond-pulse low spectral resolution images to predict picosecond-pulse high spectral resolution hyperspectral stacks. However, it is unclear how generalizable such a prediction model would be. For example, anecdotal experiments we have performed with the similar goal of spectral-denoising in the fingerprint region for drug-dosed cells suggest a loss of quantitative information after deep learning prediction. Prediction of cells dosed with drugs containing different molecular peaks were also finally predicted towards the trained drug, an

obviously undesirable feature . Further validating that deep learning predicted hyperspectral stacks reflect real observed concentration values is currently somewhat dubious. Indeed, though SRS is inherently quantitative in its signal measurement, quantification of molecules in SRS images is complicated and often tacitly implied despite insufficient validation [66]. Future work in this direction would likely need an orthogonal modality to provide a sensitive and validated concentration value with minimally cellular specificity. Quantitative mass spectrometry imaging or microfluidic assays with integrated SRS imaging may be necessary to achieve this attractive goal.

We have also shown how deep learning is valuable to enhance the analysis of the information-rich, yet complex hyperspectral images attained in SRS microcopy. Through the development of novel deep learning architectures (the UwU-Net and the modified hyperspectral ResNet) we have shown that superior segmentation, label-free prediction, and classification can be achieved even across other types of hyperspectral image datasets. While such architectures are endlessly modifiable and tunable to empirically fit a given computer vision task for hyperspectral images, we expect their value to be demonstrated in line with the canonical long-term goals of SRS: translational medicine and label-free study of fundamental biology. Towards these goals, we have shown how SRS imaging may benefit the diagnosis of thyroid cancers at the early diagnostic stage. We have also shown promising results as a method for label-free classification of T cell differentiations with active work towards reproducing such classifications on in situ images of T cells in tissue. We expect this area of research with SRS microscopy to further proliferate as deep learning methods and improved computing hardware become more and more democratized.

Finally, we are particularly excited by the long term possibilities of deep learning with label-free SRS microscopy integrated with alternate biological assays. Recently, there has been

strong interest in combining these two data spaces to achieve a wholistic picture of biological systems and molecular biology processes down to the cellular and subcellular specificity. Recent advancements of these orthogonal tools have been in parallel, with deep learning on so-called multi-omics type data showing exquisite promise and understanding the landscapes of molecular, genetic, and epigenetic variance in cell populations. Only recently, has attention turned towards the integration of biophotonics data with multi-omics data with works like OpenCell[182] or Raman2RNA[183]. The latter appears to be a particularly relevant advancement in the context of the work presented in this document, where spectral information from spontaneous Raman images is used to predict RNA-sequencing data with positive results. Though we note such work is still inchoate, utilizing only Raman information from cell nuclei (which is admittedly dubious given the unclear chemical information being mapped to a nucleotide sequence) and not using the spatial information of the Raman images, it is a promising first step in the next generation of deep learning in biophotonics.

Indeed, it is not difficult to imagine in the near future a harmonized artificial intelligence platform (as diagrammatically depicted in Figure 7.1) based on deep learning that can take as input any combination of label-free imaging data, omics-data, and potential perturbative elements (such drugs or transcription factors) and predict things like cell classification, therapeutic response, differentiation fate, cell “age”, or tissue origin location. Such a platform would be a powerful tool towards the study of biological processes, development of drugs, and development of new therapeutic methods.

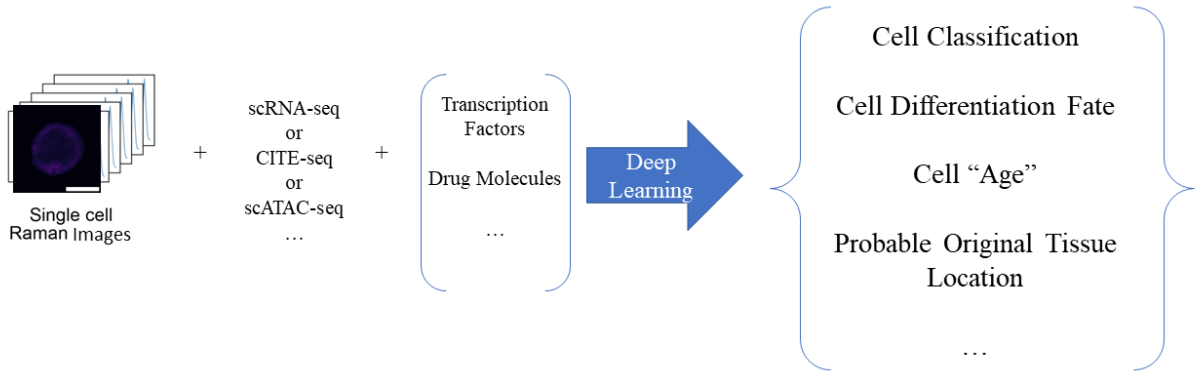

Figure 7.1. An AI platform for Biological Breakthroughs via Integration of Biophotonics and Sequencing Assays.

Overall, my contributions to this potential future platform have been established in the work of this doctoral dissertation. Through the development of deep learning's capabilities towards addressing technical shortcomings and enhancing the analysis of our chemical microscopy I have acquired a scientific training that has readied me for a continued career in pursuit of the advancement of biological knowledge at the interface of biophotonics and modern computing methods.

BIBLIOGRAPHY

1. J.-X. Cheng and X. S. Xie, "Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine," Science **350**(6264), aaa8870 (2015).
2. D. Fu, "Quantitative chemical imaging with stimulated Raman scattering microscopy," Current Opinion in Chemical Biology **39**, 24–31 (2017).
3. L. Shi, C. Zheng, Y. Shen, Z. Chen, E. S. Silveira, L. Zhang, M. Wei, C. Liu, C. de Sena-Tomas, K. Targoff, and W. Min, "Optical imaging of metabolic dynamics in animals," Nat Commun **9**, (2018).
4. D. Fu, J. Zhou, W. S. Zhu, P. W. Manley, Y. K. Wang, T. Hood, A. Wylie, and X. S. Xie, "Imaging the Intracellular Distribution of Tyrosine Kinase Inhibitors in Living Cells with Quantitative Hyperspectral Stimulated Raman Scattering," Nat Chem **6**(7), 614–622 (2014).
5. D. A. Orringer, B. Pandian, Y. S. Niknafs, T. C. Hollon, J. Boyle, S. Lewis, M. Garrard, S. L. Hervey-Jumper, H. J. L. Garton, C. O. Maher, J. A. Heth, O. Sagher, D. A. Wilkinson, M. Snuderl, S. Venneti, S. H. Ramkissoon, K. A. McFadden, A. Fisher-Hubbard, A. P. Lieberman, T. D. Johnson, X. S. Xie, J. K. Trautman, C. W. Freudiger, and S. Camelo-Piragua, "Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy," Nature Biomedical Engineering **1**(2), 0027 (2017).
6. F.-K. Lu, S. Basu, V. Igras, M. P. Hoang, M. Ji, D. Fu, G. R. Holtom, V. A. Neel, C. W. Freudiger, D. E. Fisher, and X. S. Xie, "Label-free DNA imaging in vivo with stimulated Raman scattering microscopy," Proceedings of the National Academy of Sciences **112**(37), 11624–11629 (2015).
7. B. G. Saar, C. W. Freudiger, C. M. Stanley, G. R. Holtom, and X. S. Xie, "Video-Rate Molecular Imaging In Vivo with Stimulated Raman Scattering," Science **330**(6009), 1368–1370 (2010).
8. B. Figueroa, W. Fu, T. Nguyen, K. Shin, B. Manifold, F. Wise, and D. Fu, "Broadband hyperspectral stimulated Raman scattering microscopy with a parabolic fiber amplifier source," Biomed. Opt. Express, BOE **9**(12), 6116–6131 (2018).
9. J.-X. Cheng and X. S. Xie, *Coherent Raman Scattering Microscopy* (CRC Press, 2013).
10. M. Nourhashemi, M. Mahmoudzadeh, and F. Wallois, "Thermal impact of near-infrared laser in advanced noninvasive optical brain imaging," Neurophotonics **3**(1), (2016).
11. M. L. Denton, M. S. Foltz, L. E. Estlack, D. J. Stolarski, G. D. Noojin, R. J. Thomas, D. Eikum, and B. A. Rockwell, "Damage Thresholds for Exposure to NIR and Blue Lasers in an In Vitro RPE Cell System," Invest. Ophthalmol. Vis. Sci. **47**(7), 3065–3073 (2006).
12. S. Rajan, J. Ghosh, and M. M. Crawford, "An Active Learning Approach to Hyperspectral Data Classification," IEEE Transactions on Geoscience and Remote Sensing **46**(4), 1231–1242 (2008).
13. F. Melgani and L. Bruzzone, "Support vector machines for classification of hyperspectral remote-sensing images," in *IEEE International Geoscience and Remote Sensing Symposium* (2002), **1**, pp. 506–508 vol.1.
14. W. Jahr, B. Schmid, C. Schmied, F. O. Fahrbach, and J. Huisken, "Hyperspectral light sheet microscopy," Nat Commun **6**, (2015).
15. D. Fu and X. S. Xie, "Reliable Cell Segmentation Based on Spectral Phasor Analysis of Hyperspectral Stimulated Raman Scattering Imaging Data," Anal. Chem. **86**(9), 4115–4119 (2014).

16. F. Cutrale, V. Trivedi, L. A. Trinh, C.-L. Chiu, J. M. Choi, M. S. Artiga, and S. E. Fraser, "Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging," Nat. Methods **14**(2), 149–152 (2017).
17. B.-J. Zou, Y.-D. Guo, Q. He, P.-B. Ouyang, K. Liu, and Z.-L. Chen, "3D Filtering by Block Matching and Convolutional Neural Network for Image Denoising," J. Comput. Sci. Technol. **33**(4), 838–848 (2018).
18. S. Bako, T. Vogels, B. Mcwilliams, M. Meyer, J. NováK, A. Harvill, P. Sen, T. Derose, and F. Rousselle, "Kernel-predicting Convolutional Networks for Denoising Monte Carlo Renderings," ACM Trans. Graph. **36**(4), 97:1-97:14 (2017).
19. S. K. Choudhury, P. K. Sa, R. P. Padhy, and B. Majhi, "A denoising inspired deblurring framework for regularized image restoration," in *2016 IEEE Annual India Conference (INDICON)* (2016), pp. 1–6.
20. H. Petersson, D. Gustafsson, and D. Bergstrom, "Hyperspectral image analysis using deep learning — A review," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (2016), pp. 1–6.
21. X. Ma, J. Geng, and H. Wang, "Hyperspectral image classification via contextual deep learning," J Image Video Proc. **2015**(1), 20 (2015).
22. H. Luo, "Shorten Spatial-spectral RNN with Parallel-GRU for Hyperspectral Image Classification," arXiv:1810.12563 [cs, stat] (2018).
23. X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral Image Classification With Markov Random Fields and a Convolutional Neural Network," IEEE Transactions on Image Processing **27**(5), 2354–2367 (2018).
24. S. Berisha, M. Lotfollahi, J. Jahanipour, I. Gurcan, M. Walsh, R. Bhargava, H. V. Nguyen, and D. Mayerich, "Deep learning for FTIR histology: leveraging spatial and spectral features with convolutional neural networks," Analyst **144**(5), 1642–1653 (2019).
25. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis **42**, 60–88 (2017).
26. H. Yuan, L. Cai, Z. Wang, X. Hu, S. Zhang, and S. Ji, "Computational modeling of cellular structures using conditional deep generative networks," Bioinformatics **35**(12), 2141–2149 (2019).
27. B. Manifold, E. Thomas, A. T. Francis, A. H. Hill, and D. Fu, "Denoising of stimulated Raman scattering microscopy images via deep learning," Biomed. Opt. Express, BOE **10**(8), 3860–3874 (2019).
28. A. H. Hill, A. H. Hill, B. Manifold, B. Manifold, and D. Fu, "Tissue imaging depth limit of stimulated Raman scattering microscopy," Biomed. Opt. Express, BOE **11**(2), 762–774 (2020).
29. B. Manifold, Shuaiqian Men, R. Hu, and D. Fu, "The U-within-U-Net: A Versatile Deep Learning Architecture for Classification, Segmentation, and Label-Free Prediction of Hyperspectral Images," Submitted at Nature Machine Intelligence (2020).
30. C. W. Freudiger, W. Min, B. G. Saar, S. Lu, G. R. Holtom, C. He, J. C. Tsai, J. X. Kang, and X. S. Xie, "Label-Free Biomedical Imaging with High Sensitivity by Stimulated Raman Scattering Microscopy," Science **322**(5909), 1857–1861 (2008).
31. M. J. Colles and J. E. Griffiths, "Relative and Absolute Raman Scattering Cross Sections in Liquids," J. Chem. Phys. **56**(7), 3384–3391 (1972).

32. M. O. McAnally, B. T. Phelan, R. M. Young, M. R. Wasielewski, G. C. Schatz, and R. P. Van Duyne, "Quantitative Determination of the Differential Raman Scattering Cross Sections of Glucose by Femtosecond Stimulated Raman Scattering," Anal. Chem. **89**(13), 6931–6935 (2017).
33. H. Xiong, N. Qian, Y. Miao, Z. Zhao, and W. Min, "Stimulated Raman Excited Fluorescence Spectroscopy of Visible Dyes," J. Phys. Chem. Lett. **10**(13), 3563–3570 (2019).
34. J.-X. Cheng and X. S. Xie, "Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine," Science **350**(6264), aaa8870 (2015).
35. A. H. Hill and D. Fu, "Cellular Imaging Using Stimulated Raman Scattering Microscopy," Anal. Chem. **91**(15), 9333–9342 (2019).
36. M. D. Duncan, J. Reintjes, and T. J. Manuccia, "Scanning coherent anti-Stokes Raman microscope," Opt. Lett., OL **7**(8), 350–352 (1982).
37. J.-X. Cheng and X. S. Xie, "Coherent Anti-Stokes Raman Scattering Microscopy: Instrumentation, Theory, and Applications," J. Phys. Chem. B **108**(3), 827–840 (2004).
38. D. Fu, "Quantitative chemical imaging with stimulated Raman scattering microscopy," Current Opinion in Chemical Biology **39**, 24–31 (2017).
39. A. Stender, K. Marchuk, C. Liu, S. Sander, M. Meyer, E. Smith, B. Neupane, G. Wang, J. Li, J. Cheng, B. Huang, and N. Fang, "Single Cell Optical Imaging and Spectroscopy," CHEMICAL REVIEWS **113**(4), 2469–2527 (2013).
40. I. Schie, C. Krafft, and J. Popp, "Applications of coherent Raman scattering microscopies to clinical and biological studies," ANALYST **140**(12), 3897–3909 (2015).
41. W. Min, C. W. Freudiger, S. Lu, and X. S. Xie, "Coherent nonlinear optical imaging: beyond fluorescence microscopy," Annu Rev Phys Chem **62**, 507–530 (2011).
42. C. H. Camp Jr and M. T. Cicerone, "Chemically sensitive bioimaging with coherent Raman scattering," Nat Photon **9**(5), 295–305 (2015).
43. F. Masia, A. Karuna, P. Borri, and W. Langbein, "Hyperspectral image analysis for CARS, SRS, and Raman data," Journal of Raman Spectroscopy **46**(8), 727–734 (2015).
44. P. Nandakumar, A. Kovalev, and A. Volkmer, "Vibrational imaging based on stimulated Raman scattering microscopy," NEW JOURNAL OF PHYSICS **11**, (2009).
45. W. Min, C. W. Freudiger, S. Lu, and X. S. Xie, "Coherent Nonlinear Optical Imaging: Beyond Fluorescence Microscopy," Annu. Rev. Phys. Chem. **62**(1), 507–530 (2011).
46. M. J. Sanderson, I. Smith, I. Parker, and M. D. Bootman, "Fluorescence Microscopy," Cold Spring Harb Protoc **2014**(10), pdb.top071795 (2014).
47. P. A. Santi, "Light Sheet Fluorescence Microscopy: A Review," J Histochem Cytochem. **59**(2), 129–138 (2011).
48. S. Ranjit, L. Lanzanò, A. E. Libby, E. Gratton, and M. Levi, "Advances in fluorescence microscopy techniques to study kidney function," Nat Rev Nephrol **17**(2), 128–144 (2021).
49. J. Icha, M. Weber, J. C. Waters, and C. Norden, "Phototoxicity in live fluorescence microscopy, and how to avoid it," BioEssays **39**(8), 1700003 (2017).
50. K. Sepp, M. Lee, M. T. J. Bluntzer, G. V. Helgason, A. N. Hulme, and V. G. Brunton, "Utilizing Stimulated Raman Scattering Microscopy To Study Intracellular Distribution of Label-Free Ponatinib in Live Cells," J. Med. Chem. **63**(5), 2028–2034 (2020).
51. A. Feizpour, T. Marstrand, L. Bastholm, S. Eirefelt, and C. L. Evans, "Label-Free Quantification of Pharmacokinetics in Skin with Stimulated Raman Scattering Microscopy and Deep Learning," Journal of Investigative Dermatology **141**(2), 395–403 (2021).

52. C.-S. Liao, M. N. Slipchenko, P. Wang, J. Li, S.-Y. Lee, R. A. Oglesbee, and J.-X. Cheng, "Microsecond scale vibrational spectroscopic imaging by multiplex stimulated Raman scattering microscopy," Light: Science & Applications **4**(3), e265 (2015).
53. K. Bae, W. Zheng, Y. Ma, and Z. Huang, "Real-Time Monitoring of Pharmacokinetics of Mitochondria-Targeting Molecules in Live Cells with Bioorthogonal Hyperspectral Stimulated Raman Scattering Microscopy," Anal. Chem. **92**(1), 740–748 (2020).
54. N. Belsey, N. Garrett, L. Contreras-Rojas, A. Pickup-Gerlaugh, G. Price, J. Moger, and R. Guy, "Evaluation of drug delivery to intact and porated skin by coherent Raman scattering and fluorescence microscopies," JOURNAL OF CONTROLLED RELEASE **174**, 37–42 (2014).
55. C. Stiebing, T. Meyer, I. Rimke, C. Matthäus, M. Schmitt, S. Lorkowski, and J. Popp, "Real-time Raman and SRS imaging of living human macrophages reveals cell-to-cell heterogeneity and dynamics of lipid uptake," Journal of Biophotonics **10**(9), 1217–1226 (2017).
56. L. Zhang, L. Shi, Y. Shen, Y. Miao, M. Wei, N. Qian, Y. Liu, and W. Min, "Spectral tracing of deuterium for imaging glucose metabolism," Nat Biomed Eng **3**(5), 402–413 (2019).
57. C. Zhang, J. Li, L. Lan, and J.-X. Cheng, "Quantification of Lipid Metabolism in Living Cells through the Dynamics of Lipid Droplets Measured by Stimulated Raman Scattering Imaging," Anal. Chem. **89**(8), 4502–4507 (2017).
58. K.-C. Huang, J. Li, C. Zhang, Y. Tan, and J.-X. Cheng, "Multiplex Stimulated Raman Scattering Imaging Cytometry Reveals Lipid-Rich Protrusions in Cancer Cells under Stress Condition," iScience **23**(3), 100953 (2020).
59. M. Wang, W. Min, C. Freudiger, G. Ruvkun, and X. Xie, "RNAi screening for fat regulatory genes with SRS microscopy," NATURE METHODS **8**(2), 135-U52 (2011).
60. W. Hong, C. W. Karanja, N. S. Abutaleb, W. Younis, X. Zhang, M. N. Seleem, and J.-X. Cheng, "Antibiotic Susceptibility Determination within One Cell Cycle at Single-Bacterium Level by Stimulated Raman Metabolic Imaging," Anal. Chem. **90**(6), 3737–3743 (2018).
61. K. Bae, W. Zheng, Y. Ma, and Z. Huang, "Real-Time Monitoring of Pharmacokinetics of Antibiotics in Biofilms with Raman-Tagged Hyperspectral Stimulated Raman Scattering Microscopy," THERANOSTICS **9**(5), 1348–1357 (2019).
62. B. Zhang, H. Xu, J. Chen, X. Zhu, Y. Xue, Y. Yang, J. Ao, Y. Hua, and M. Ji, "Highly specific and label-free histological identification of microcrystals in fresh human gout tissues with stimulated Raman scattering," THERANOSTICS **11**(7), 3074–3088 (2021).
63. L. Zhang, Y. Wu, B. Zheng, L. Su, Y. Chen, S. Ma, Q. Hu, X. Zou, L. Yao, Y. Yang, L. Chen, Y. Mao, Y. Chen, and M. Ji, "Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy," Theranostics **9**(9), 2541–2554 (2019).
64. M. Ji, M. Arbel, L. Zhang, C. W. Freudiger, S. S. Hou, D. Lin, X. Yang, B. J. Bacskai, and X. S. Xie, "Label-free imaging of amyloid plaques in Alzheimer's disease with stimulated Raman scattering microscopy," Science Advances **4**(11), eaat7715 (2018).
65. Y. Yang, Y. Yang, Z. Liu, L. Guo, S. Li, X. Sun, Z. Shao, and M. Ji, "Microcalcification-Based Tumor Malignancy Evaluation in Fresh Breast Biopsies with Hyperspectral Stimulated Raman Scattering," Anal. Chem. **93**(15), 6223–6231 (2021).
66. B. Manifold and D. Fu, "Quantitative Stimulated Raman Scattering Microscopy: Promises and Pitfalls," Annual Review of Analytical Chemistry **15**(1), null (2022).

67. A. H. Hill, E. Munger, A. T. Francis, B. Manifold, and D. Fu, "Frequency Modulation Stimulated Raman Scattering Microscopy through Polarization Encoding," J. Phys. Chem. B **123**(40), 8397–8404 (2019).
68. M. Wei, L. Shi, Y. Shen, Z. Zhao, A. Guzman, L. J. Kaufman, L. Wei, and W. Min, "Volumetric chemical imaging by clearing-enhanced stimulated Raman scattering microscopy," PNAS **116**(14), 6608–6617 (2019).
69. X. Zhang, M. B. J. Roeffaers, S. Basu, J. R. Daniele, D. Fu, C. W. Freudiger, G. R. Holtom, and X. S. Xie, "Label-Free Live-Cell Imaging of Nucleic Acids Using Stimulated Raman Scattering Microscopy," ChemPhysChem **13**(4), 1054–1059 (2012).
70. S. Adabi, S. Ghavami, M. Fatemi, and A. Alizad, "Non-Local Based Denoising Framework for In Vivo Contrast-Free Ultrasound Microvessel Imaging," Sensors (Basel) **19**(2), (2019).
71. S. Chitchian, M. A. Fiddy, and N. M. Fried, "Denoising during optical coherence tomography of the prostate nerves via wavelet shrinkage using dual-tree complex wavelet transform," J Biomed Opt **14**(1), 014031 (2009).
72. E. A. Pnevmatikakis, D. Soudry, Y. Gao, T. A. Machado, J. Merel, D. Pfau, T. Reardon, Y. Mu, C. Lacefield, W. Yang, M. Ahrens, R. Bruno, T. M. Jessell, D. S. Peterka, R. Yuste, and L. Paninski, "Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data," Neuron **89**(2), 285–299 (2016).
73. G. Gilboa, Y. Y. Zeevi, and N. Sochen, "Signal and image enhancement by a generalized forward-and-backward adaptive diffusion process," in *2000 10th European Signal Processing Conference* (2000), pp. 1–4.
74. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(7), 629–639 (1990).
75. A. Buades, B. Coll, and J.- Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (2005), **2**, pp. 60–65 vol. 2.
76. D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang, "When Image Denoising Meets High-Level Vision Tasks: A Deep Learning Approach," (2017).
77. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," Nature Methods **15**(11), 917–920 (2018).
78. B. Figueroa, Y. Chen, K. Berry, A. Francis, and D. Fu, "Label-Free Chemical Imaging of Latent Fingerprints with Stimulated Raman Scattering Microscopy," Anal. Chem. **89**(8), 4468–4473 (2017).
79. A. Francis, K. Berry, Y. Chen, B. Figueroa, and D. Fu, "Label-free pathology by spectrally sliced femtosecond stimulated Raman scattering (SRS) microscopy," PLoS ONE **12**(5), e0178750 (2017).
80. A. T. Francis, T. T. Nguyen, M. S. Lamm, R. Teller, S. P. Forster, W. Xu, T. Rhodes, R. L. Smith, J. Kuiper, Y. Su, and D. Fu, "In Situ Stimulated Raman Scattering (SRS) Microscopy Study of the Dissolution of Sustained-Release Implant Formulation," Mol. Pharm. **15**(12), 5793–5801 (2018).
81. D. Fu, G. Holtom, C. Freudiger, X. Zhang, and X. S. Xie, "Hyperspectral Imaging with Stimulated Raman Scattering by Chirped Femtosecond Lasers," J. Phys. Chem. B **117**(16), 4634–4640 (2013).
82. V. Chauhan, P. Bowlan, J. Cohen, and R. Trebino, "Single-diffraction-grating and grism pulse compressors," J. Opt. Soc. Am. B, JOSAB **27**(4), 619–624 (2010).

83. R. He, Y. Xu, L. Zhang, S. Ma, X. Wang, D. Ye, and M. Ji, "Dual-phase stimulated Raman scattering microscopy for real-time two-color imaging," Optica, OPTICA **4**(1), 44–47 (2017).
84. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds., Lecture Notes in Computer Science (Springer International Publishing, 2015), pp. 234–241.
85. C. Zhang, D. Zhang, and J.-X. Cheng, "Coherent Raman Scattering Microscopy in Biology and Medicine," Annu. Rev. Biomed. Eng. **17**(1), 415–445 (2015).
86. L. Azzari and A. Foi, "Variance stabilization in Poisson image deblurring," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (2017), pp. 728–731.
87. D. Sage and M. Unser, "Teaching image-processing programming in Java," IEEE Signal Processing Magazine **20**(6), 43–52 (2003).
88. J. Li, F. Luisier, and T. Blu, "Deconvolution of poissonian images with the PURE-LET approach," in *2016 IEEE International Conference on Image Processing (ICIP)* (2016), pp. 2708–2712.
89. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds., Lecture Notes in Computer Science (Springer International Publishing, 2015), pp. 234–241.
90. M. Ji, S. Lewis, S. Camelo-Piragua, S. H. Ramkissoon, M. Snuderl, S. Venneti, A. Fisher-Hubbard, M. Garrard, D. Fu, A. C. Wang, J. A. Heth, C. O. Maher, N. Sanai, T. D. Johnson, C. W. Freudiger, O. Sagher, X. S. Xie, and D. A. Orringer, "Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy," Sci Transl Med **7**(309), 309ra163 (2015).
91. L. Zhang, Y. Wu, B. Zheng, L. Su, Y. Chen, S. Ma, Q. Hu, X. Zou, L. Yao, Y. Yang, L. Chen, Y. Mao, Y. Chen, and M. Ji, "Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy," Theranostics **9**(9), 2541–2554 (2019).
92. P. Coupé, M. Munz, J. V. Manjón, E. S. Ruthazer, and D. Louis Collins, "A CANDLE for a deeper in vivo insight," Medical Image Analysis **16**(4), 849–864 (2012).
93. H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," IEEE Transactions on Computational Imaging **3**(1), 47–57 (2017).
94. A. H. Hill, B. Manifold, and D. Fu, "Tissue imaging depth limit of stimulated Raman scattering microscopy," Biomed. Opt. Express **11**(2), 762 (2020).
95. A. J. Wright, S. P. Poland, J. M. Girkin, C. W. Freudiger, C. L. Evans, and X. S. Xie, "Adaptive optics for enhanced signal in CARS microscopy," Opt. Express, OE **15**(26), 18209–18219 (2007).
96. C. L. Evans, E. O. Potma, M. Puoris'haag, D. Côté, C. P. Lin, and X. S. Xie, "Chemical imaging of tissue in vivo with video-rate coherent anti-Stokes Raman scattering microscopy," PNAS **102**(46), 16807–16812 (2005).
97. H. Wang, T. B. Huff, Y. Fu, K. Y. Jia, and J.-X. Cheng, "Increasing the imaging depth of coherent anti-Stokes Raman scattering microscopy with a miniature microscope objective," Opt. Lett., OL **32**(15), 2212–2214 (2007).

98. D. M. Drutis, T. M. Hancewicz, E. Pashkovski, L. Feng, D. Mihalov, G. R. Holtom, K. P. Ananthapadmanabhan, X. S. Xie, and M. Misra, "Three-dimensional chemical imaging of skin using stimulated Raman scattering microscopy," JBO **19**(11), 111604 (2014).
99. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods **15**(12), 1090–1097 (2018).
100. X. Mao, C. Shen, and Y.-B. Yang, "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2016), **29**.
101. F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," Nat Methods **2**(12), 932–940 (2005).
102. M. Wang, C. Wu, D. Sinefeld, B. Li, F. Xia, and C. Xu, "Comparing the effective attenuation lengths for long wavelength in vivo imaging of the mouse brain," Biomed. Opt. Express, BOE **9**(8), 3534–3543 (2018).
103. E. Chaigneau, A. J. Wright, S. P. Poland, J. M. Girkin, and R. A. Silver, "Impact of wavefront distortion and scattering on 2-photon microscopy in mammalian brain tissue," Opt. Express, OE **19**(23), 22755–22774 (2011).
104. B. G. Saar, C. W. Freudiger, C. M. Stanley, G. R. Holtom, and X. S. Xie, "Video-Rate Molecular Imaging In Vivo with Stimulated Raman Scattering," Science **330**(6009), 1368–1370 (2010).
105. M. J. B. Moester, L. Zada, B. Fokker, F. Ariese, and J. F. de Boer, "Stimulated Raman scattering microscopy with long wavelengths for improved imaging depth," Journal of Raman Spectroscopy **50**(9), 1321–1328 (2019).
106. J. Li, P. Lin, Y. Tan, J.-X. Cheng, J.-X. Cheng, and J.-X. Cheng, "Volumetric stimulated Raman scattering imaging of cleared tissues towards three-dimensional chemical histopathology," Biomed. Opt. Express, BOE **10**(8), 4329–4339 (2019).
107. D. S. Richardson and J. W. Lichtman, "Clarifying Tissue Clearing," Cell **162**(2), 246–257 (2015).
108. Y. Chen, S. Liu, H. Liu, S. Tong, H. Tang, C. Zhang, S. Yan, H. Li, G. Yang, D. Zhu, K. Wang, and P. Wang, "Coherent Raman Scattering Unravelling Mechanisms Underlying Skull Optical Clearing for Through-Skull Brain Imaging," Anal. Chem. **91**(15), 9371–9375 (2019).
109. Y.-J. Zhao, T.-T. Yu, C. Zhang, Z. Li, Q.-M. Luo, T.-H. Xu, and D. Zhu, "Skull optical clearing window for *in vivo* imaging of the mouse cortex at synaptic resolution," Light: Science & Applications **7**(2), 17153 (2018).
110. B. Manifold, S. Men, R. Hu, and D. Fu, "A versatile deep learning architecture for classification and label-free prediction of hyperspectral images," Nat Mach Intell **3**(4), 306–315 (2021).
111. E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine **25**(1), 44–56 (2019).
112. S. Mittal, C. Stoean, A. Kajdacsy-Balla, and R. Bhargava, "Digital Assessment of Stained Breast Tissue Images for Comprehensive Tumor and Microenvironment Analysis," Front. Bioeng. Biotechnol. **7**, (2019).

113. P. Mukherjee, M. Zhou, E. Lee, A. Schicht, Y. Balagurunathan, S. Napel, R. Gillies, S. Wong, A. Thieme, A. Leung, and O. Gevaert, "A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets," Nature Machine Intelligence **2**(5), 274–282 (2020).
114. O. S. Pianykh, S. Guitron, D. Parke, C. Zhang, P. Pandharipande, J. Brink, and D. Rosenthal, "Improving healthcare operations management with machine learning," Nature Machine Intelligence **2**(5), 266–273 (2020).
115. M. Varma, M. Lu, R. Gardner, J. Dunnmon, N. Khandwala, P. Rajpurkar, J. Long, C. Beaulieu, K. Shpanskaya, L. Fei-Fei, M. P. Lungren, and B. N. Patel, "Automated abnormality detection in lower extremity radiographs using deep learning," Nature Machine Intelligence **1**(12), 578–583 (2019).
116. L. Zhang, Y. Wu, B. Zheng, L. Su, Y. Chen, S. Ma, Q. Hu, X. Zou, L. Yao, Y. Yang, L. Chen, Y. Mao, Y. Chen, and M. Ji, "Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy," Theranostics **9**(9), 2541–2554 (2019).
117. A. Rana, A. Lowe, M. Lithgow, K. Horback, T. Janovitz, A. Da Silva, H. Tsai, V. Shanmugam, A. Bayat, and P. Shah, "Use of Deep Learning to Develop and Analyze Computational Hematoxylin and Eosin Staining of Prostate Core Biopsy Images for Tumor Diagnosis," JAMA Netw Open **3**(5), (2020).
118. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images," Cell **173**(3), 792-803.e19 (2018).
119. Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," IEEE Transactions on Geoscience and Remote Sensing **49**(10), 3973–3985 (2011).
120. K. Klein, A. M. Gigler, T. Aschenbrenner, R. Monetti, W. Bunk, F. Jamitzky, G. Morfill, R. W. Stark, and J. Schlegel, "Label-Free Live-Cell Imaging with Confocal Raman Microscopy," Biophysical Journal **102**(2), 360–368 (2012).
121. L. Mou, P. Ghamisi, and X. X. Zhu, "Deep Recurrent Neural Networks for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing **55**(7), 3639–3655 (2017).
122. S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep Learning for Hyperspectral Image Classification: An Overview," IEEE Transactions on Geoscience and Remote Sensing **57**(9), 6690–6709 (2019).
123. Xiang Li, W. Li, Xiaodong Xu, and Wei Hu, "Cell classification using convolutional neural networks in medical hyperspectral imagery," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)* (2017), pp. 501–504.
124. W. Zhao, Z. Guo, J. Yue, X. Zhang, and L. Luo, "On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery," International Journal of Remote Sensing **36**(13), 3368–3379 (2015).
125. D. Mayerich, M. J. Walsh, A. Kadjacsy-Balla, P. S. Ray, S. M. Hewitt, and R. Bhargava, "Stain-less staining for computed histopathology," Technology **03**(01), 27–31 (2015).
126. J. Behrmann, C. Etmann, T. Boskamp, R. Casadonte, J. Kriegsmann, and P. Maaβ, "Deep learning for tumor classification in imaging mass spectrometry," Bioinformatics **34**(7), 1215–1223 (2018).

127. J. Zhang, J. Zhao, H. Lin, Y. Tan, and J.-X. Cheng, "High-Speed Chemical Imaging by Dense-Net Learning of Femtosecond Stimulated Raman Scattering," J. Phys. Chem. Lett. **11**(20), 8573–8578 (2020).
128. T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, A. Dovzhenko, O. Tietz, C. D. Bosco, S. Walsh, D. Saltukoglu, T. L. Tay, M. Prinz, K. Palme, M. Simons, I. Diester, T. Brox, and O. Ronneberger, "U-Net: deep learning for cell counting, detection, and morphometry," Nature Methods 1 (2018).
129. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," Nature Methods **15**(11), 917–920 (2018).
130. F. Foroozandeh Shahraki, L. Saadatifard, S. Berisha, M. Lotfollahi, D. Mayerich, and S. Prasad, "Deep Learning for Hyperspectral Image Analysis, Part II: Applications to Remote and Biomedicine," in (2020), pp. 69–115.
131. A. Soni, R. Koner, and V. G. K. Villuri, "M-UNet: Modified U-Net Segmentation Framework with Satellite Imagery," in *Proceedings of the Global AI Congress 2019*, J. K. Mandal and S. Mukhopadhyay, eds., Advances in Intelligent Systems and Computing (Springer, 2020), pp. 47–59.
132. B. Cui, Y. Zhang, X. Li, J. Wu, and Y. Lu, "WetlandNet: Semantic Segmentation for Remote Sensing Images of Coastal Wetlands via Improved UNet with Deconvolution," in *Genetic and Evolutionary Computing*, J.-S. Pan, J. C.-W. Lin, Y. Liang, and S.-C. Chu, eds., Advances in Intelligent Systems and Computing (Springer, 2020), pp. 281–292.
133. N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention Unet for building segmentation in remote sensing images," Sci. China Inf. Sci. **63**(4), 140305 (2020).
134. M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3," (n.d.).
135. Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks," IEEE Transactions on Geoscience and Remote Sensing **54**(10), 6232–6251 (2016).
136. Z. Meng, L. Li, X. Tang, Z. Feng, L. Jiao, and M. Liang, "Multipath Residual Network for Spectral-Spatial Hyperspectral Image Classification," Remote Sensing **11**(16), 1896 (2019).
137. Z. Xue, "A general generative adversarial capsule network for hyperspectral image spectral-spatial classification," Remote Sensing Letters **11**(1), 19–28 (2020).
138. S. A. Thomas, Y. Jin, J. Bunch, and I. S. Gilmore, "Enhancing classification of mass spectrometry imaging data with deep neural networks," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (2017), pp. 1–8.
139. T. Alexandrov and A. Bartels, "Testing for presence of known and unknown molecules in imaging mass spectrometry," Bioinformatics **29**(18), 2335–2342 (2013).
140. F. Tobias, J. C. McIntosh, G. J. LaBonia, M. W. Boyce, M. R. Lockett, and A. B. Hummon, "Developing a Drug Screening Platform: MALDI-Mass Spectrometry Imaging of Paper-Based Cultures," Anal. Chem. **91**(24), 15370–15376 (2019).
141. C. D. Wijetunge, I. Saeed, B. A. Boughton, J. M. Spraggins, R. M. Caprioli, A. Bacic, U. Roessner, and S. K. Halgamuge, "EXIMS: an improved data analysis pipeline based on a new peak picking method for EXploring Imaging Mass Spectrometry data," Bioinformatics **31**(19), 3198–3206 (2015).
142. A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, M. Becker, and T. Alexandrov, "FDR-controlled

metabolite annotation for high-resolution imaging mass spectrometry," Nat. Methods **14**(1), 57–60 (2017).
143. X. Liu, J. K. Lukowski, C. Flinders, S. Kim, R. A. Georgiadis, S. M. Mumenthaler, and A. B. Hummon, "MALDI-MSI of Immunotherapy: Mapping the EGFR-Targeting Antibody Cetuximab in 3D Colon-Cancer Cell Cultures," Anal. Chem. **90**(24), 14156–14164 (2018).
144. P. Inglese, G. Correia, Z. Takats, J. K. Nicholson, and R. C. Glen, "SPUTNIK: an R package for filtering of spatially related peaks in mass spectrometry imaging data," Bioinformatics **35**(1), 178–180 (2019).
145. J. M. Fonville, C. L. Carter, L. Pizarro, R. T. Steven, A. D. Palmer, R. L. Griffiths, P. F. Lalor, J. C. Lindon, J. K. Nicholson, E. Holmes, and J. Bunch, "Hyperspectral Visualization of Mass Spectrometry Imaging Data," Anal. Chem. **85**(3), 1415–1423 (2013).
146. P. Inglese, J. S. McKenzie, A. Mroz, J. Kinross, K. Veselkov, E. Holmes, Z. Takats, J. K. Nicholson, and R. C. Glen, "Deep learning and 3D-DESI imaging reveal the hidden metabolic heterogeneity of cancer," Chemical Science **8**(5), 3500–3511 (2017).
147. D. Fu, G. Holtom, C. Freudiger, X. Zhang, and X. S. Xie, "Hyperspectral Imaging with Stimulated Raman Scattering by Chirped Femtosecond Lasers," J. Phys. Chem. B **117**(16), 4634–4640 (2013).
148. T. A. Pologruto, B. L. Sabatini, and K. Svoboda, "ScanImage: Flexible software for operating laser scanning microscopes," BioMed Eng OnLine **2**(1), 13 (2003).
149. J. O. Eriksson, M. Rezeli, M. Hefner, G. Marko-Varga, and P. Horvatovich, "Clusterwise Peak Detection and Filtering Based on Spatial Distribution To Efficiently Mine Mass Spectrometry Imaging Data," Anal. Chem. **91**(18), 11888–11896 (2019).
150. L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," IEEE Transactions on Image Processing **20**(8), 2378–2386 (2011).
151. A. Santara, K. Mani, P. Hatwar, A. Singh, A. Garg, K. Padia, and P. Mitra, "BASS Net: Band-Adaptive Spectral-Spatial Feature Learning Neural Network for Hyperspectral Image Classification," IEEE Trans. Geosci. Remote Sensing **55**(9), 5293–5301 (2017).
152. S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification," IEEE Geosci. Remote Sensing Lett. **17**(2), 277–281 (2020).
153. Y. Liang, X. Zhao, A. J. X. Guo, and F. Zhu, "Hyperspectral Image Classification With Deep Metric Learning and Conditional Random Field," IEEE Geoscience and Remote Sensing Letters **17**(6), 1042–1046 (2020).
154. B. M. Prentice, C. W. Chumbley, and R. M. Caprioli, "High-Speed MALDI MS/MS Imaging Mass Spectrometry Using Continuous Raster Sampling," J Mass Spectrom **50**(4), 703–710 (2015).
155. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 5967–5976.
156. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nature Biomedical Engineering **3**(6), 466–477 (2019).
157. C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," Nature Methods **16**(12), 1215–1225 (2019).

158. D. Sage and M. Unser, "Teaching image-processing programming in Java," IEEE Signal Processing Magazine **20**(6), 43–52 (2003).
159. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.
160. N. Bayramoglu, M. Kaakinen, L. Eklund, and J. Heikkilä, "Towards Virtual H E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (2017), pp. 64–71.
161. S. M. Kaech, E. J. Wherry, and R. Ahmed, "Effector and memory T-cell differentiation: implications for vaccine development," Nat Rev Immunol **2**(4), 251–262 (2002).
162. M. H. Kershaw, J. A. Westwood, and P. K. Darcy, "Gene-engineered T cells for cancer therapy," Nat Rev Cancer **13**(8), 525–541 (2013).
163. J. M. Chu, N. A. Pease, and H. Y. Kueh, "In search of lost time: Enhancers as modulators of timing in lymphocyte development and differentiation," Immunological Reviews **300**(1), 134–151 (2021).
164. K. Abadie, N. A. Pease, M. J. Wither, and H. Y. Kueh, "Order by chance: origins and benefits of stochasticity in immune cell fate control," Current Opinion in Systems Biology **18**, 95–103 (2019).
165. J. Y. M. Lee and P. E. Love, "Assessment of T Cell Development by Flow Cytometry," in *T-Cell Development: Methods and Protocols*, R. Bosselut and M. S. Vacchio, eds., Methods in Molecular Biology (Springer, 2016), pp. 47–64.
166. J. O. Ross, H. J. Melichar, J. Halkias, and E. A. Robey, "Studying T Cell Development in Thymic Slices," in *T-Cell Development: Methods and Protocols*, R. Bosselut and M. S. Vacchio, eds., Methods in Molecular Biology (Springer, 2016), pp. 131–140.
167. L. K. Lutes, Z. Steier, L. L. McIntyre, S. Pandey, J. Kaminski, A. R. Hoover, S. Ariotti, A. Streets, N. Yosef, and E. A. Robey, "T cell self-reactivity during thymic development dictates the timing of positive selection," eLife **10**, e65435 (2021).
168. M. P. Matheu, I. Parker, and M. D. Cahalan, "Dissection and 2-Photon Imaging of Peripheral Lymph Nodes in Mice," JoVE (Journal of Visualized Experiments) (7), e265 (2007).
169. "Pytorch Cats and Dogs Classification," https://kaggle.com/adinishad/pytorch-cats-and-dogs-classification.
170. H. Wu, V. Estrella, M. Beatty, D. Abrahams, A. El-Kenawi, S. Russell, A. Ibrahim-Hashim, D. L. Longo, Y. K. Reshetnyak, A. Moshnikova, O. A. Andreev, K. Luddy, M. Damaghi, K. Kodumudi, S. R. Pillai, P. Enriquez-Navas, S. Pilon-Thomas, P. Swietach, and R. J. Gillies, "T-cells produce acidic niches in lymph nodes to suppress their own effector functions," Nat Commun **11**(1), 4113 (2020).
171. J. M. Leal, J. Y. Huang, K. Kohli, C. Stoltzfus, M. R. Lyons-Cohen, B. E. Olin, M. Gale, and M. Y. Gerner, "Innate cell microenvironments in lymph nodes shape the generation of T cell responses during type I inflammation," Science Immunology **6**(56), eabb9435 (2021).
172. C. R. Stoltzfus, J. Filipek, B. H. Gern, B. E. Olin, J. M. Leal, Y. Wu, M. R. Lyons-Cohen, J. Y. Huang, C. L. Paz-Stoltzfus, C. R. Plumlee, T. Pöschinger, K. B. Urdahl, M. Perro, and M. Y. Gerner, "CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues," Cell Reports **31**(3), (2020).

173. R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," CA: A Cancer Journal for Clinicians **72**(1), 7–33 (2022).
174. E. S. Cibas and S. Z. Ali, "The 2017 Bethesda System for Reporting Thyroid Cytopathology," Thyroid **27**(11), 1341–1346 (2017).
175. L. Hegedüs, "The Thyroid Nodule," New England Journal of Medicine **351**(17), 1764–1771 (2004).
176. Z. W. Baloch, V. A. LiVolsi, S. L. Asa, J. Rosai, M. J. Merino, G. Randolph, P. Vielh, R. M. DeMay, M. K. Sidawy, and W. J. Frable, "Diagnostic terminology and morphologic criteria for cytologic diagnosis of thyroid lesions: A synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference," Diagnostic Cytopathology **36**(6), 425–437 (2008).
177. R. L. Ferris, Z. Baloch, V. Bernet, A. Chen, T. J. Fahey, I. Ganly, S. P. Hodak, E. Kebebew, K. N. Patel, A. Shaha, D. L. Steward, R. P. Tufano, S. M. Wiseman, and S. E. Carty, "American Thyroid Association Statement on Surgical Application of Molecular Profiling for Thyroid Nodules: Current Impact on Perioperative Decision Making," Thyroid **25**(7), 760–768 (2015).
178. M. Halicek, M. Halicek, J. D. Dormer, J. V. Little, A. Y. Chen, B. Fei, and B. Fei, "Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning," Biomed. Opt. Express, BOE **11**(3), 1383–1400 (2020).
179. M. A. S. de Oliveira, M. Campbell, A. M. Afify, E. C. Huang, E. C. Huang, J. W. Chan, and J. W. Chan, "Hyperspectral Raman microscopy can accurately differentiate single cells of different human thyroid nodules," Biomed. Opt. Express, BOE **10**(9), 4411–4421 (2019).
180. M. A. S. de Oliveira, M. Campbell, A. M. Afify, E. C. Huang, E. C. Huang, J. W. Chan, and J. W. Chan, "Raman-based cytopathology: an approach to improve diagnostic accuracy in medullary thyroid carcinoma," Biomed. Opt. Express, BOE **11**(12), 6962–6972 (2020).
181. "Instant diagnosis of gastroscopic biopsy via deep-learned single-shot femtosecond stimulated Raman histology," https://www.researchsquare.com.
182. N. H. Cho, K. C. Cheveralls, A.-D. Brunner, K. Kim, A. C. Michaelis, P. Raghavan, H. Kobayashi, L. Savy, J. Y. Li, H. Canaj, J. Y. S. Kim, E. M. Stewart, C. Gnann, F. McCarthy, J. P. Cabrera, R. M. Brunetti, B. B. Chhun, G. Dingle, M. Y. Hein, B. Huang, S. B. Mehta, J. S. Weissman, R. Gómez-Sjöberg, D. N. Itzhak, L. A. Royer, M. Mann, and M. D. Leonetti, "OpenCell: Endogenous tagging for the cartography of human cellular organization," Science **375**(6585), eabi6983 (n.d.).
183. K. J. Kobayashi-Kirschvink, S. Gaddam, T. James-Sorenson, E. Grody, J. R. Ounadjela, B. Ge, K. Zhang, J. W. Kang, R. Xavier, P. T. C. So, T. Biancalani, J. Shu, and A. Regev, "Raman2RNA: Live-cell label-free prediction of single-cell RNA expression profiles by Raman microscopy," 2021.11.30.470655 (2021).

APPENDIX A

This Appendix contains additional figures for:

Chapter 2: Deep learning Denoising of SRS Microscopy Images

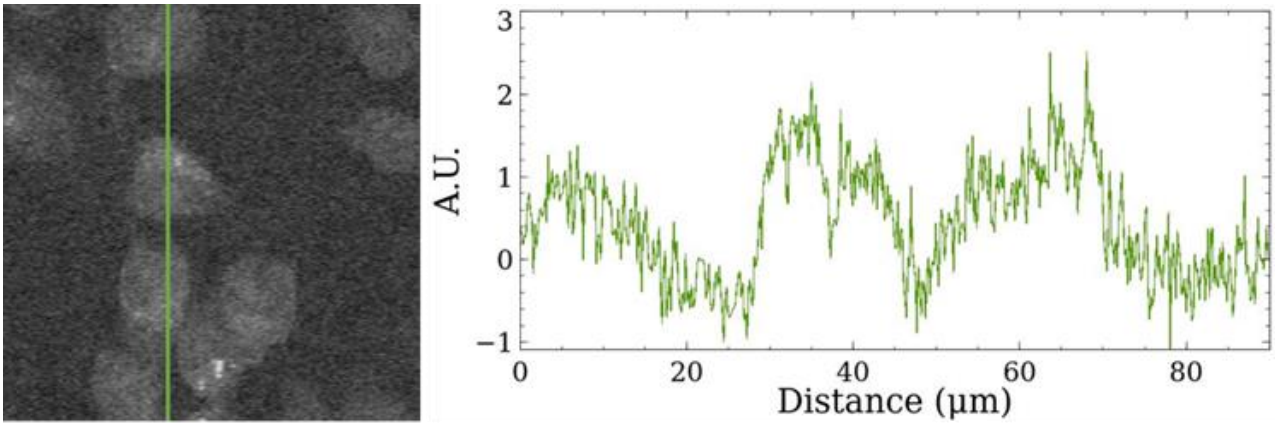

Figure A.A.1. Image of HeLa Cells Shown in Figure 2.2 Denoised using PURE-LET While cells are slightly denoised and separated from the background, the variance of the pixel values remains largely the same, and spatial features are significantly blurred together.

Eq. (A.A.1) shows how PSNR is calculated where *r* and *t* are the reference (truth) and tested (denoised) image respectively, (x,y) is a given pixel coordinate in an n_x x n_y image, and *R* is the maximum value of the tested image. Images that contain more noise will have lower PSNR values.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{R^2}{\sum_{r,t}[r(x,y)-t(x,y)]^2 / n_x * n_y}\right) \tag{A.A.1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_x n_y}\cdot\sum_{0}^{n_x-1}\sum_{0}^{n_y-1}\left[r(x,y)-t(x,y)\right]^2} \tag{A.A.2}$$

Eq. (A.A.2) shows how RMSE is calculated. The RMSE is a value that expresses the accuracy of a given denoising method based on a calculation of error at each pixel coordinate between the two images. RMSE values closer to 0 indicate a more accurate denoising. Note that PSNR implicitly utilizes the RMSE squared or mean square error.

$$\mathrm{CC} = \frac{\sum_{0}^{n_x-1}\sum_{0}^{n_y-1}\left[r(x,y)-\bar{r}\right]\left[t(x,y)-\bar{t}\right]}{\sqrt{\sum_{0}^{n_x-1}\sum_{0}^{n_y-1}\left[r(x,y)-\bar{r}\right]^2}\sqrt{\sum_{0}^{n_x-1}\sum_{0}^{n_y-1}\left[t(x,y)-\bar{t}\right]^2}} \tag{A.A.3}$$

Eq. (A.A.3) shows how CC (also known as the Pearson correlation coefficient) is calculated where $\bar{r}$ and $\bar{t}$ are the average pixel value for the reference (truth) and test (denoised) image respectively. The CC is a measure of the covariance between the reference and test image divided by the product of the standard deviations of the respective images. Here, if features are localized with one another between two images the CC trends towards 1. Noise in the test image results in CC values closer to 0.

APPENDIX B

This Appendix contains additional figures for:

Chapter 3: Extending the Imaging Depth Limit of SRS Microscopy

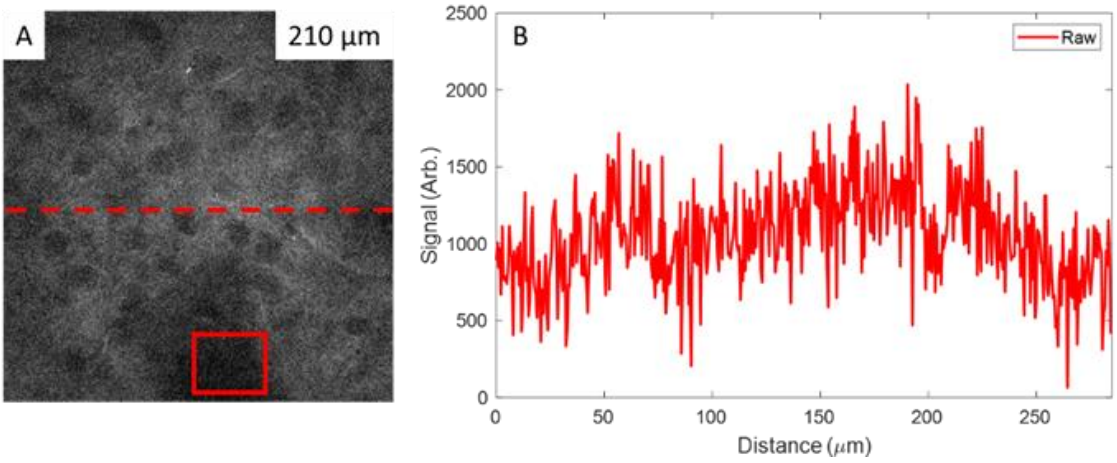

Figure A.B.1. Raw SRS Image of Murine Brain at 210 µm Deep

(A) Raw image collected at 210 µm deep in the murine cortex. (B) Line plot of area depicted by dashed line in panel A. Background used in SBR calculation collected from area encompassed by red square.

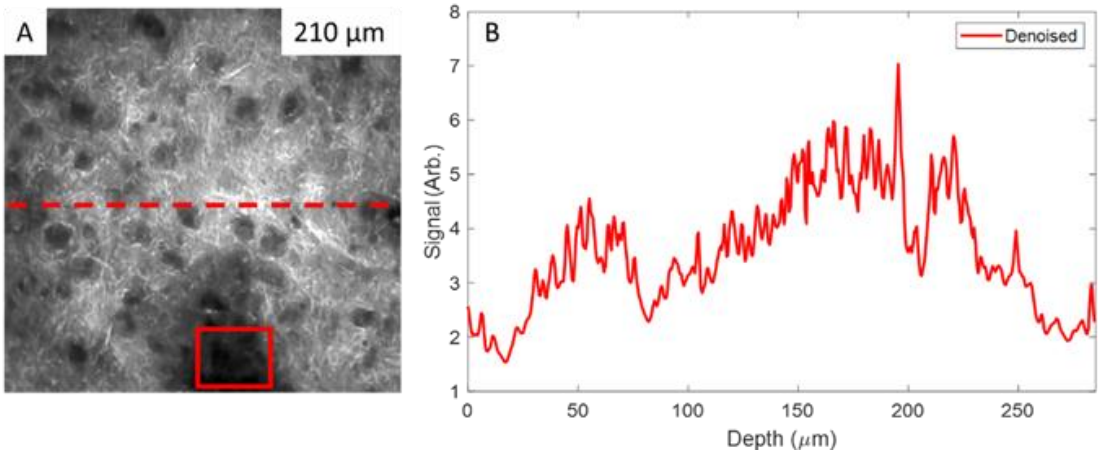

Figure A.B. 2. Deep Learning Denoised SRS Image of Murine Brain at 210 µm Deep

(A) Raw image collected at 210 μm deep in the murine cortex. (B) Line plot of area depicted by dashed line in panel A. Background used in SBR calculation collected from area encompassed by red square.

APPENDIX C

This Appendix contains additional figures for:

Chapter 4: The U-within-U-Net A Versatile Deep Learning Architecture for Classification and Label-Free Prediction of Hyperspectral Images

Table A.C.1. False Pixels Counts and Intersection Over Union in the UwU-Net (17-U) Indian Pines Model

Classification Label	False Positive Pixels	False Negative Pixels	IOU
Alfalfa	1	8	0.836
Corn (No Till)	65	231	0.803
Corn (Min Till)	19	149	0.806
Corn	11	35	0.812
Grass (Pasture)	6	101	0.787
Grass (Trees)	9	32	0.948
Grass (Mowed Pasture)	3	2	0.828
Hay (Windrowed)	13	6	0.962
Oats	0	4	0.800
Soybeans (No Till)	59	112	0.834
Soybeans (Min Till)	210	103	0.883
Soybeans (Clean Till)	80	79	0.769
Wheat	4	10	0.935
Woods	75	96	0.875
Buildings	9	161	0.753
Stone-Steel Tower	6	13	0.812
Overall	570	1142	0.840

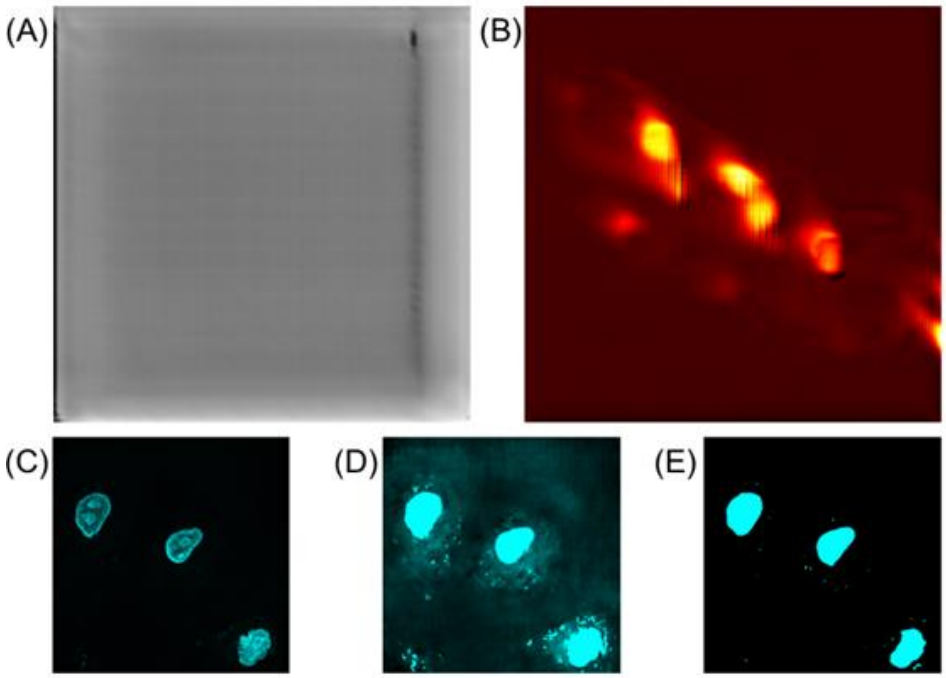

Figure A.C. 1. Representative Predictions from Facile U-Net for Hyperspectral Images Panel a shows the Grass (Mowed Pasture) Indian Pines classification prediction with no thresholding. Panel b shows the prediction of Ipratropium from the MSI dataset. Panel c shows prediction of nuclear fluorescence from SRS images with contrast values set to mimic the images shown in Figure 4.3. Panel d shows the same image as Panel c with higher contrast to demonstrate the U-Net's inability to remove non-nucleus features. Panel e shows the UwU-Net prediction from Figure 4.3a with high contrast demonstrating superior non-nuclear feature removal.

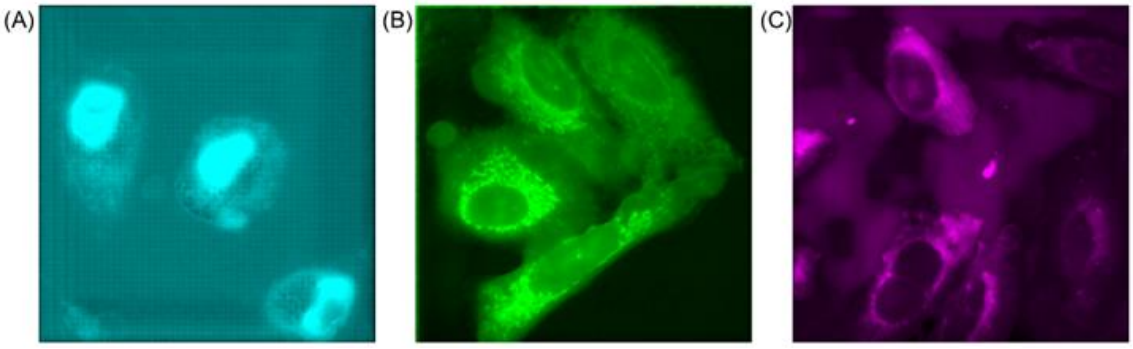

Figure A.C. 2. Fluorescence Predictions from the Modified U-Net with with ResNet Blocks Panel a shows nucleus fluorescence prediction. Panel b shows mitochondrial prediction. Panel c shows endoplasmic reticulum prediction. All truth fields of view are the same as in Figure 4.3.

Table A.C. 2. Quality Metrics for Res-U-Net

Organelle Model	PCC	NRMSE	FSIM
Nucleus	0.74 ± 0.04	0.379 ± 0.016	0.76 ± 0.05
Mitochondria	0.75 ± 0.15	0.172 ± 0.053	0.77 ± 0.06
Endoplasmic Reticulum	0.72 ± 0.13	0.112 ± 0.023	0.78 ± 0.03

PCC, NRMSE, and FSIM values for the Res-U-Net trained as in Extended Data Figure 2. The number of images for used for each calculation is the same as in Table 4.3. Uncertainty refers to standard deviation.

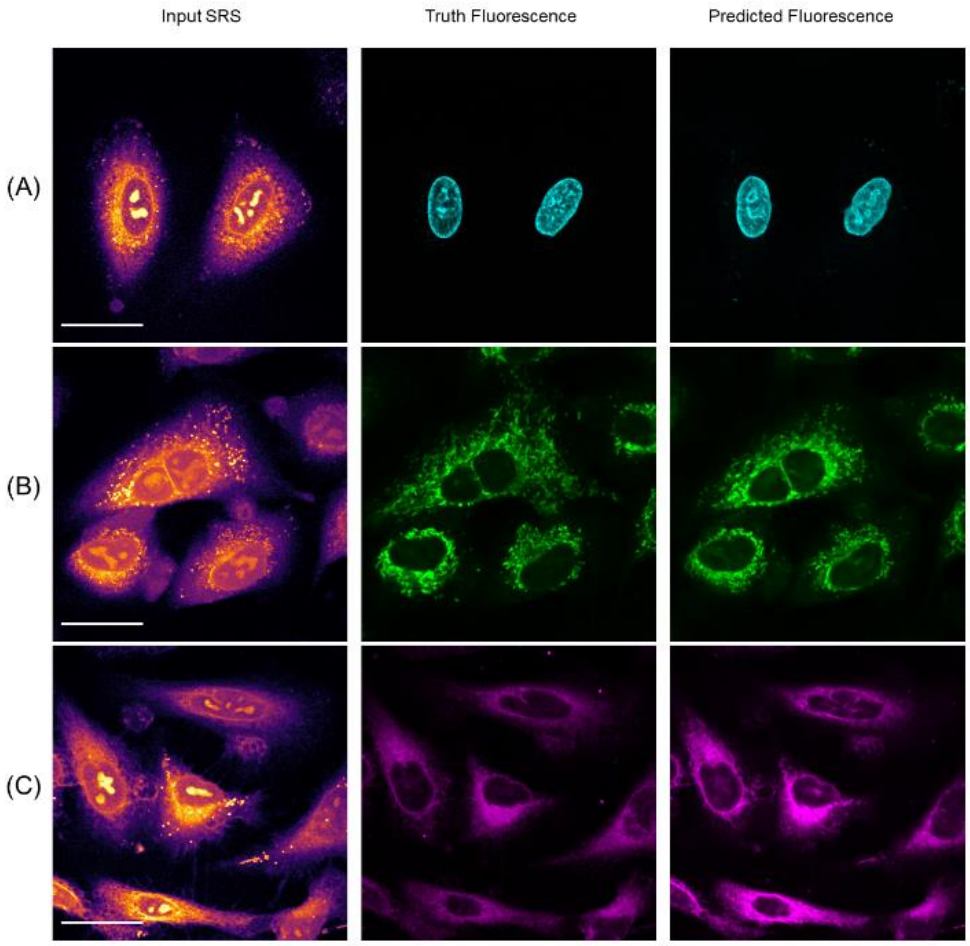

Figure A.C. 3. Predicted Organelle fluorescence using traditional U-Net

Panel a shows prediction of nucleus fluorescence. Panel b shows prediction of mitochondrial fluorescence. Panel c shows prediction of endoplasmic reticulum fluorescence. We note the improper inclusion of lipid droplets in the mitochondria model and off nucleoli in both the mitochondria and endoplasmic reticulum models. The comparison between lipid droplets and mitochondria is further depicted in Figure A.C.4.

Table A.C. 3. Quality Metrics for Traditional U-Net Fluorescence Prediction

Organelle Model	PCC
Nucleus	0.84 ± 0.05
Mitochondria	0.81 ± 0.05
Endoplasmic Reticulum	0.93 ± 0.03

PCC metrics for the organelle fluorescence prediction models trained with a traditional U-Net using a single SRS image. While still highly correlated, we note the errant prediction of spurious features in Figures A.C.3 and A.C.4.

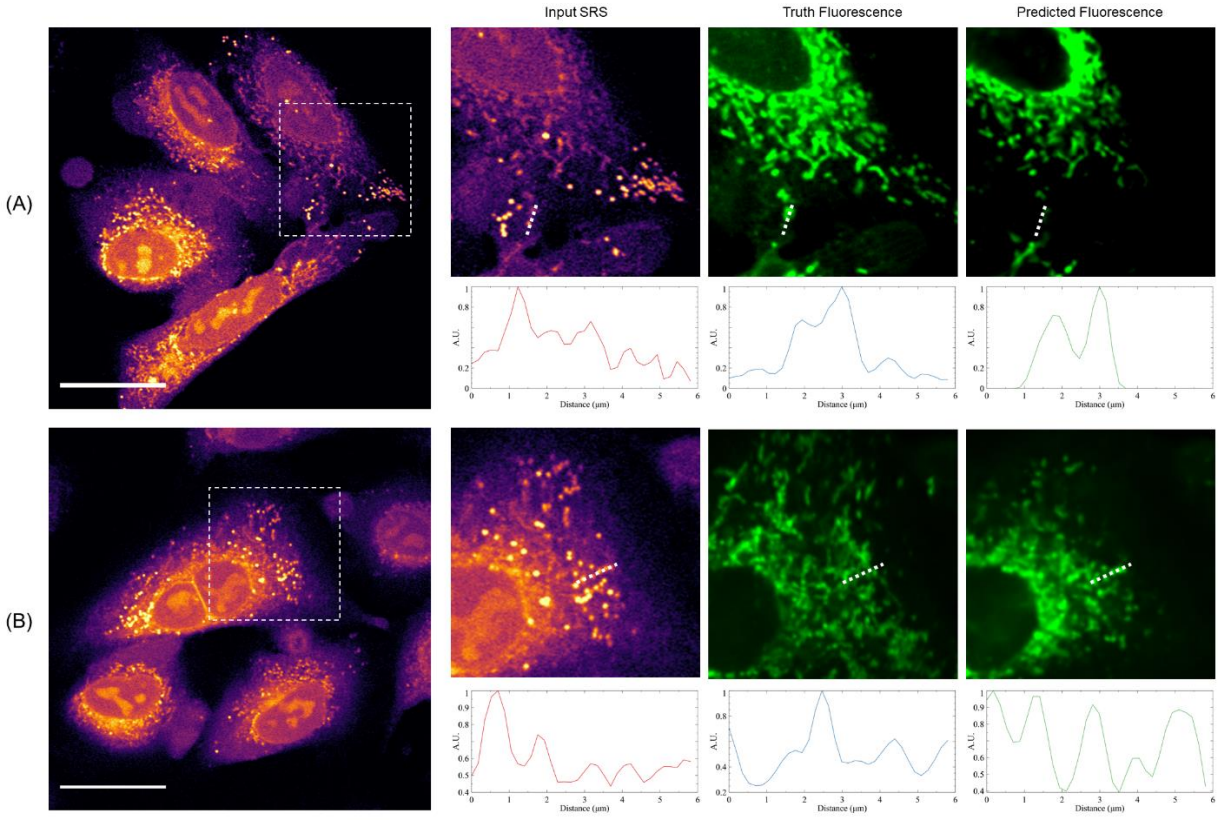

Figure A.C. 4. Comparison of mitochondria prediction between UwU-Net and traditional U-Net.

Panel a shows a zoomed in field of view from Figure 1b where a UwU-Net is trained to predict mitochondrial fluorescence from a hyperspectral SRS stack. The shown input SRS only corresponds to the brightest image out of the 10-image hyperspectral stack. Normalized pixel values are plotted below each image corresponding to the drawn dashed lines. In the SRS image, a strong lipid droplet is found at ~1.4 μm but is properly removed during prediction of the mitochondria at ~1.8 μm and ~3 μm. Panel b shows a zoomed in field of view from Extended Data Figure 3b where a traditional U-Net is trained to predict mitochondrial fluorescence from a single SRS image. The normalized pixel value plots beneath each zoomed-in field of view show a marked difference in how lipid droplets are handled. Here the lipid droplets at ~0.8 μm and ~1.8 μm are not removed during prediction.

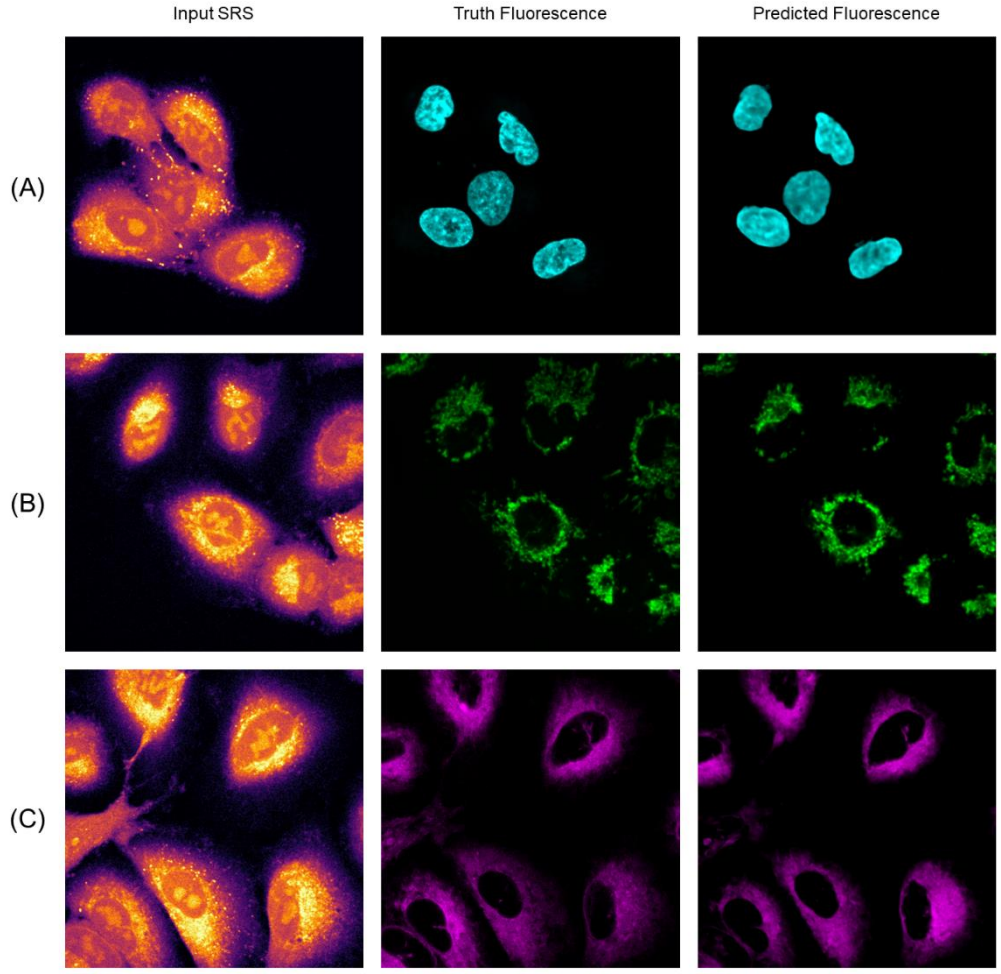

Figure A.C. 5. UwU-Net predicted fluorescence in live-cell SRS imaging.

Panel a shows prediction of nucleus fluorescence. Panel b shows prediction of mitochondrial fluorescence. Panel c shows prediction of endoplasmic reticulum fluorescence.

BRYCE ADRIAN MANIFOLD

Brycemanifold8@gmail.com · github.com/B-Manifold · (317) 697-9791

Objective

Experienced microscopist with well-cited publications creating and utilizing deep learning models for biomedical image analysis. Seeking to leverage technical skills and knowledge at the interface of artificial intelligence and real-world problems in highly interdisciplinary cutting-edge research.

Education

University of Washington, Seattle, WA July 2017 – June 2022
Ph.D. in Chemistry
GPA: 3.90

Indiana University, Bloomington, IN August 2013 – May 2017
Bachelor of Science in Chemistry with High Distinction and Honors
Minors: Mathematics and Music
GPA: 3.88

Professional Research Experience

Dan Fu Research Group – Seattle, WA July 2017 – Present
Graduate Researcher

- Developed deep learning algorithms and models to improve analysis of Stimulated Raman Scattering microscopy of cancer cells and brain tissue
- Built and maintained ultrafast laser microscopes for stimulated Raman Scattering, multiphoton fluorescence, transient absorption, and second harmonic generation microscopies
- Published multiple peer reviewed articles as both a lead and supporting author in reputable scientific journals
- Presented award-winning work at the largest international Photonics conference

Megan Thielges Research Group – Bloomington, IN August 2013 – May 2017
Undergraduate Researcher

- Utilized Fourier Transform Infrared and Ultraviolet-Visible Spectroscopy to elucidate the mechanisms of substrate affinity in cytochrome P450cam
- Published results in *Biochemistry* as a supporting author
- Presented published results at annual IU Chemistry Awards Banquet

Procter & Gamble – Cincinnati, OH May – August 2014 & 2015
Research and Development Intern

- Leveraged skills in team coordination, analytical chemistry, and data analysis to optimize production of fabric care and home-cleaning products
- Presented weekly findings and final outcomes of research project to senior engineers, international partners, and technical management

Lead Author Publications

B Manifold; D Fu; Deep Learning Diagnosis of Thyroid Cancer from Hyperspectral Stimulated Raman Scattering Microscopy Images. *In preparation,* **(2022)**

B Manifold; D Fu; Label-Free in situ Classification of Immune Cells from Hyperspectral Stimulated Raman Scattering Microscopy Images. *In preparation,* **(2022)**

B Manifold; D Fu; Quantitative Stimulated Raman Scattering Microscopy – Promises and Pitfalls. *Annu. Rev. Anal. Chem.* **(2022)**

B Manifold; S Men; R Hu; D Fu; A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. *Nature Machine Intelligence,* **(2021)**

B Manifold; B Figueroa; D Fu; Stimulated Raman Scattering Microscopy Chapter 4: Hyperspectral SRS imaging via spectral focusing. *Elsevier, ISBN: 978-0-323-85158-9*, **(2021)**

AH Hill*; **B Manifold***; D Fu; Tissue Imaging Depth Limit of Stimulated Raman Scattering Microscopy. *Biomedical Optics Express*, **(2020)** (*co-first author)

B Manifold; EC Thomas; AT Francis; AH Hill; D Fu; Denoising of Stimulated Raman Scattering Microscopy Images via Deep Learning. *Biomedical Optics Express,* **(2019)**

Supporting Author Publications

AT Francis; **B Manifold**; EC Thomas; R Hu; AH Hill; S Men; D Fu; In Vivo Simultaneous Nonlinear Absorption Raman and Fluorescence (SNARF) Imaging of Mouse Brain Cortical Structures. *Commun. Biol.,* **(2022)**

AH Hill; E Munger; AT Francis; **B Manifold**; D Fu; Frequency modulation stimulated Raman scattering microscopy through polarization encoding. *The Journal of Physical Chemistry B,* **(2019)**

B Figueroa; W Fu; T Nguyen; K Shin; **B Manifold**; F Wise; D Fu; Broadband hyperspectral stimulated Raman scattering microscopy with a parabolic fiber amplifier source. *Biomed. Opt. Express*, **(2018)**

EJ Basom; **BA Manifold**; MC Thielges; Conformational Heterogeneity and the Affinity of Substrate Molecular Recognition by Cytochrome P450cam. *Biochemistry*, **(2017)**

Technical Skills

Programming and Image Analysis

- Python
- Pytorch
- MATLAB
- Bash
- Deep Learning
- Machine Learning
- ImageJ
- ImageJ Macro

Optical System Design and Implementation

- SRS Microscopy
- Multiphoton Fluorescence Microscopy
- Ultrafast Laser Systems
- Pulse Shaping
- Fiber Optic Splicing
- Fiber Laser Amplification
- Supercontinuum Generation
- Optical Parametric Oscillators
- High Frequency Modulation
- Lock-in Amplification

Awards and Honors

SPIE Photonics Europe Grant	March 2022
Martin P. Gouterman Endowed Award	March 2021
SPIE Photonics West Presentation Award	February 2020
National Science Foundation GRFP Honorable Mention	March 2018
Lloyd and Florence West Fellowship	September 2017
Excellence in Chemistry Graduate Fellowship	September 2017
Early Bird Research Assistantship	July 2017
Viola Scholarship in Nuclear Chemistry	May 2017
Indiana University Chemistry Departmental Honors	May 2017
Indiana University Chemistry Honor roll	August 2013 – May 2017
Indiana University Founders Scholar	August 2013 – May 2017
IU College of Arts & Sciences Executive Dean's List	August 2013 – May 2017
Phi Beta Kappa Honor Society Inductee	December 2016
Robert & Marjorie Mann Research Scholarship	May 2016
Forrest L Warner Chemistry Scholarship	May 2016
Indiana University Academic Excellence Scholar	August 2013
IU College of Arts & Science Direct Admission Scholar	August 2013

Additional Professional Experience

Music and Arts and SoundSync Studios – Redmond, WA August 2017 – Present

Cello Teacher

- Maintained a studio of 10 students with ages ranging 12 to 50 years old
- Taught cello methodology and basic music theory to beginner through early advanced students
- Performed with fellow music teachers to engage new students and community in music

Residential Programs and Services – Bloomington, IN August 2014 – May 2017

Resident Assistant

- Fostered a diverse community of residents focusing on academic development leading to significantly increased community GPA's
- Created and maintained a residential community aligned with Indiana University's model of residential life
- Served the Indiana University and Bloomington community through didactic programing

IU All-Campus and UW Campus Philharmonic Orchestras January 2015 – August 2019

Principal Cellist

- Led the cello section during rehearsals and performances
- Conducted and taught sectional rehearsals for cellists
- Collaborated with the director and other section leaders to standardize performance elements