# Virtual Staining of Photon Absorption Remote Sensing Histology Images: A Promising Approach for Label-Free Histological Imaging

by

Marian Boktor

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2023

© Marian Boktor 2023

**Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

| | |
|---|---|
| External Examiner: | Chulhong Kim<br>Professor, Dept. of Artificial Intelligence and Interdisciplinary Bioscience and Bioengineering,<br>Pohang University of Science and Technology |
| Supervisor(s): | Parsin Haji Reza<br>Associate Professor, Dept. of Systems Design Engineering,<br>University of Waterloo<br><br>Paul Fieguth<br>Professor, Dept. of Systems Design Engineering,<br>University of Waterloo |
| Internal Member(s): | Maud Gorbet<br>Professor, Dept. of Systems Design Engineering,<br>University of Waterloo<br><br>Alex Wong<br>Professor, Dept. of Systems Design Engineering,<br>University of Waterloo |
| Internal-External Member: | Karim Sallaudin Karim<br>Professor, Dept. of Electrical & Computer Engineering,<br>University of Waterloo |

## Author's Declaration

I hereby declare that this thesis consists of material all of which I either authored or co-authored: see Statement of Contributions. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Statement of Contributions

Marian Boktor was the sole author of Chapters 1, 2, 4, 6, and 7.

Chapter 3 of this thesis consists of material from the published paper [1], co-authored with Benjamin R. Ecclestone, Vlad Pekar, Deepak Dinakaran, John R. Mackey, Paul Fieguth, and Parsin Haji Reza.

Chapter 5 of this thesis consists of material from the paper under review [2], co-authored with James E.D. Tweel, Benjamin R. Ecclestone, Jennifer Ai Ye, Paul Fieguth, and Parsin Haji Reza.

All the datasets used in this study were collected by Benjamin R. Ecclestone and James E.D. Tweel.

## Abstract

Histology is a crucial diagnostic tool in medicine and biological research, providing valuable insights into tissue morphology and cellular structures. However, conventional histological staining methods are time-consuming. In recent years, label-free microscopy such as photon absorption remote sensing (PARS) has emerged as a promising alternative. PARS microscope utilizes the optical absorption properties of target absorbers to capture high-resolution images of tissues at various wavelengths. This thesis focuses on the development of virtual staining pipelines for PARS histology images.

First, a hematoxylin and eosin (H&E) virtual staining pipeline is introduced utilizing state-of-the-art image-to-image translation models, namely pix2pix and cycleGAN. The pipeline consists of preprocessing, image registration, model training, and model testing stages. The trained models facilitate the translation of PARS images into stained versions, replicating the appearance of H&E-stained histology images. The results demonstrate a strong concordance between virtually stained images and true H&E, with pathologists reviewing the results of one of the studies and confirming that the diagnostic quality of the virtually stained PARS images closely resembles that of gold standard H&E.

While PARS has made substantial progress in virtual staining, certain challenges persist, particularly regarding model confusion between tissue structures, which results in colorization ambiguity. To enhance structure distinction, we propose expanding the number of input channels in the virtual staining models. This involves introducing a multi-channel virtual staining pipeline that includes feature extraction from PARS time-resolved signals using an existing modified K-means algorithm and a proposed feature selection method based on metrics like the Structural Similarity Index (SSIM). This pipeline aims to enhance the virtual staining process through feature labeling prior to model training. The results show that optimized feature combinations outperform conventional PARS channels in virtually staining human skin and mouse brain tissue, yielding a significant visual and quantitative agreement with gold-standard H&E images.

Finally, the proposed virtual staining frameworks are modified to accommodate a broader range of stains and tissue types. In order to showcase the versatility of the frameworks, we apply them to datasets containing Periodic Acid-Schiff (PAS) stain in addition to H&E, specifically using a human fungus tissue sample. The experimental results show that incorporating additional features outperforms conventional PARS channels in both H&E and PAS stains, suggesting the potential of this approach to enhance tissue structure labeling and improve different staining techniques. Moreover, this approach demonstrates the feasibility of using PARS to generate different virtual stains from the same data sources.

Throughout the thesis, extensive experimentation and evaluation are conducted using diverse PARS datasets and ground truth histology images. Performance metrics, such as accuracy and visual fidelity, in addition to clinical validation, are employed to assess the effectiveness of the virtual staining pipelines. The findings presented in this thesis demonstrate the efficacy and versatility of the proposed approaches, contributing to the advancement of slide-free histological imaging, which promises to dramatically reduce the time from specimen resection to histological imaging.

**Acknowledgements**

I would like to extend my heartfelt appreciation to my esteemed supervisors, Professor Paul Fieguth and Professor Parsin Haji Reza, for their unwavering support, invaluable guidance, and endless patience throughout this research journey. Their mentorship and encouragement have consistently pushed me to strive for excellence and become a better researcher. Their insightful feedback and constructive criticism have significantly improved my writing and presentation skills. This research owes its success to their profound enthusiasm, wisdom, and valuable suggestions, and I am truly honored to have had the opportunity to work under their guidance. I am deeply grateful for their dedication and belief in my potential.

I am immensely grateful to my colleagues at the PhotoMedicine Labs for fostering an exceptional work atmosphere, encouraging scientific discussions, and providing valuable insights. Special thanks go to Ben, James, and Alex for their insightful contributions to our discussions. Ben and James made substantial contributions, particularly in the collection of all datasets employed in this thesis. Moreover, I extend my heartfelt appreciation to Lyazzat, Tara, Layla, Nima, Kris, Hager, and Stephanie, who have not only been outstanding colleagues but also supportive friends, making the journey more enjoyable and fulfilling. Lastly, I am sincerely thankful to Nick and Vlad for always being available to offer help, engage in discussions, and provide support, which has been invaluable throughout this endeavor. I would also like to convey my appreciation to my master's advisor, Dr. Ahmad, for providing valuable support, both emotionally and technically, during the challenging moments brought about by the pandemic.

I want to express my gratitude to my examiners, Professor Wong, Professor Gorbet, Professor Karim, and Professor Kim for their meticulous review of the thesis, as well as for their constructive feedback and valuable suggestions.

I would also like to express my genuine gratitude to my family abroad, whose unwavering support has been a constant source of strength and motivation for me. Their love and encouragement have carried me through the ups and downs of this journey. Additionally, I am deeply grateful to my chosen family in Canada, especially Philo, whose presence and enduring emotional support have been invaluable during the challenging moments and setbacks. Their belief in me and their encouragement has been instrumental in helping me persevere and achieve my goals. I cannot thank them enough for being a pillar of strength in my life.

This thesis received continuous support from multiple clinicians serving as a valuable resource throughout the research. I would like to extend my gratitude to Dr. Marie Abi

Daoud at the Alberta Precision Laboratories in Calgary, Canada, and the Cross-Cancer Institute in Edmonton, Alberta, for providing human skin tissue samples and Dr. Deepak Dinakaran and Dr. Kevin Camphausen from the radiation oncology branch at the National Cancer Institute, NIH, Bethesda, MD, USA for providing mouse brain samples.

I am also thankful to Dr. Gilbert Bigras, Dr. Marie Abi Daoud, Dr. Charlene Hunter, and Dr. Karen Naert for their expert pathology review of some of the PARS images. Additionally, I acknowledge the valuable assistance of Hager Gaouda in preparing and staining the majority of the tissue samples used in this thesis.

Finally, I gratefully acknowledge the financial support provided by the following funding sources throughout the duration of this thesis: Natural Sciences and Engineering Research Council of Canada (DGECR-2019-00143, RGPIN201906134, DH-2023-00371), Canada Foundation for Innovation (JELF #38000), Mitacs Accelerate (IT13594), University of Waterloo Startup funds, Centre for Bioengineering and Biotechnology (CBB Seed fund), illumiSonics Inc (SRA #083181), New frontiers in research fund – exploration (NFRFE-2019-01012), The Canadian Institutes of Health Research (CIHR PJT 185984), and Discovery Horizons DH-2023-00371.

## Dedication

This is dedicated to cancer patients all over the world.

# Table of Contents

**Examining Committee Membership** ii

**Author's Declaration** iii

**Statement of Contributions** iv

**Abstract** v

**Acknowledgements** vii

**Dedication** ix

**List of Figures** xiii

**List of Tables** xvi

**1 Introduction** 1

**2 Preliminaries** 5

2.1 Background . . . . . 5

2.1.1 Histopathology . . . . . 5

2.1.2 Histological Staining Techniques . . . . . 7

2.1.3 Label-Free Microscopy . . . . . 9

2.1.4 PARS . . . 10
2.1.5 Virtual Staining . . . 23
2.1.6 Generative Adversarial Networks (GANs) . . . 24
2.1.7 Modified K-means Clustering (K*-means) . . . 28
2.2 Related Work . . . 31

**3 PARS-to-H&E Translation using pix2pix GANs 52**
3.1 Introduction . . . 52
3.2 Dataset Preparation . . . 53
3.3 PARS Imaging . . . 55
3.4 Dataset Preparation for Training . . . 55
3.5 Implementation Details . . . 60
3.6 Clinical Validation of Virtual H&E Images of Skin . . . 60
3.7 Results and Discussion . . . 61
3.8 Conclusion . . . 67

**4 A Comparison between Pix2pix GAN and CycleGAN for PARS-to-H&E Translation 70**
4.1 Introduction . . . 70
4.2 Methods . . . 71
4.2.1 Sample Preparation . . . 71
4.2.2 Sample Imaging and Image Formation . . . 72
4.2.3 Image Registration . . . 73
4.2.4 Implementation Details . . . 74
4.3 Results and Discussion . . . 76
4.4 Conclusion . . . 81

**5 Multi-Channel Feature Extraction for PARS-to-H&E Translation** **82**
5.1 Motivation . . . 82
5.2 Introduction . . . 84
5.3 Sample Preparation and Data Acquisition . . . 86
5.4 PARS Imaging . . . 87
5.5 Image Registration and Preprocessing . . . 88
5.6 Time-Domain Feature Extraction . . . 90
5.6.1 K*-means Clustering . . . 90
5.6.2 Principal Component Analysis (PCA) . . . 91
5.7 Multi-Channel GAN (MC-GAN) . . . 93
5.8 Training Settings . . . 94
5.9 Results and Discussion . . . 94
5.9.1 Feature Extraction . . . 95
5.9.2 Feature Selection . . . 103
5.9.3 Virtual Staining Results . . . 103
5.10 Conclusion . . . 112

**6 Application: PARS-to-PAS Translation of Human Skin Fungus** **115**
6.1 Introduction . . . 115
6.2 Dataset Preparation . . . 116
6.3 Training Settings . . . 116
6.4 Results and Discussion . . . 117
6.5 Conclusion . . . 125

**7 Conclusions and Future Work** **126**

**References** **128**

# List of Figures


2.1 Images of H&E and PAS stains of human skin fungus tissue. . . . . 8

2.2 Absorption spectra of the major endogenous chromophores of biological tissue. Source: Reprinted from [3]. . . . . 13

2.3 Different implementations of PAI and their penetration limit. . . . . 17

2.4 The typical setup of PARS system. . . . . 21

2.5 Overview of a GAN training algorithm. Source: Reprinted from [4]. . . . . 24

2.6 The architectures of two image translation models: (a) pix2pix and (b) cycleGAN. . . . . 29

2.7 The results of digital staining using hyperspectral images. . . . . 35

2.8 Experimental verification of the proposed network using four datasets of bright-field images. . . . . 37

2.9 The StainGAN results were compared with the cycleGAN and pix2pix models. 38

2.10 A comparison between MUSE, Deep-MUSE, and H&E images. . . . . 40

2.11 The results of pseudo-H&E virtual staining of RCM images. . . . . 42

2.12 Molecular imaging using digital staining methods with varied spatial resolutions. . . . . 44

2.13 Virtual H&E histology generation from OCT images. Source: Reprinted from [5]. . . . . 46

2.14 Virtual staining of unlabeled autofluorescence images. . . . . 47

2.15 The validation results of UV-PAM and Deep-PAM. . . . . 50

3.1 Virtual staining of PARS images via deep learning. . . . . 54

3.2 Simplified PARS system architecture and system setup. . . . . 56
3.3 Virtual staining framework. . . . . 58
3.4 The colorization network input and output images of human skin tissue slide. 59
3.5 An illustration of the pix2pix training algorithm. . . . . 61
3.6 Virtual staining of PARS images of human skin tissue slides. . . . . 63
3.7 Virtual staining results of different parts of human skin tissue. . . . . 64
3.8 Another set of virtual staining results using human skin tissue. . . . . 66
3.9 A comparison between virtual staining outputs using different inputs. . . . 68

4.1 MATLAB GUI used for registration. . . . . 75
4.2 The overlay displays the image registration results of PARS and H&E images. 76
4.3 Comparison of cycleGAN and pix2pix virtual staining on human skin. . . . 78
4.4 Comparison of cycleGAN (paired and unpaired) and pix2pix virtual staining on human skin. . . . . 79

5.1 Enhancing colorization with separated H&E channels. . . . . 85
5.2 Simplified PARS histology optical architecture. . . . . 89
5.3 K*-means feature extraction enhances tissue differentiation. . . . . 92
5.4 Multi-channel virtual staining workflow. . . . . 96
5.5 Feature extraction results of the K-study. . . . . 97
5.6 Virtual staining results of the K-study for each $K \in \{2, \ldots, 6\}$. . . . . 98
5.7 Feature extraction results using PCA. . . . . 100
5.8 Virtual Staining results using PCA. . . . . 101
5.9 Example of PARS channels using human skin. . . . . 105
5.10 Comparing PARS feature combinations for virtual staining. . . . . 107
5.11 Analyzing PARS feature combinations for virtual staining. . . . . 109
5.12 Comparing virtual staining: K*-means ($K = 2$, $K = 3$) vs. conventional channels. . . . . 111
5.13 Another comparison of virtual staining results of mouse brain tissue. . . . . 113

6.1 Virtual PAS staining of PARS images using conventional channels. . . . . 118
6.2 An illustration of various image domains of human skin tissue. . . . . 120
6.3 Virtual staining: conventional vs. best-performing channels. . . . . 122
6.4 Enhancing virtual PAS staining with multiple channels. . . . . 124

# List of Tables

2.1 Summary of GAN-based frameworks. . . . . 33

4.1 Quantitative analysis of H&E colorization using cycleGAN, pix2pix (loose), pix2pix (tight). . . . . 80

5.1 Ablation study for feature extraction. . . . . 102

5.2 Summary of quantitative analysis of the C-study. . . . . 106

5.3 Ablation study for the best feature combination. . . . . 108

6.1 Quantitative analysis summary of H&E and PAS colorization results. . . . 121

# Chapter 1

# Introduction

Histopathology serves as a fundamental microscopic examination tool across diverse domains, encompassing cancer diagnosis and prognosis, surgical oncology, and drug discovery and development [6]. Traditional histological staining techniques, such as hematoxylin and eosin (H&E) staining, have been the gold standard for visualizing tissue structures and identifying cellular components.

This conventional procedure involves multiple steps [7, 8], starting with the fixation of the tissue sample using a fixative, commonly formalin, to preserve its cellular structure and prevent decay. After fixation, the tissue is embedded in paraffin wax, providing support and facilitating the sectioning process. Sectioning, accomplished with a microtome, refers to slicing the embedded tissue into thin sections, typically around 5 to 10 micrometers thick. These thin sections are then mounted on glass slides and subjected to various staining techniques to enhance the visibility of cellular structures and specific molecules of interest.

While this approach has been indispensable for decades, providing detailed tissue visualizations, it suffers from several limitations. The laborious and time-consuming nature of the process hinders its application in urgent clinical situations [7] where rapid diagnosis is vital. Additionally, the use of chemical fixatives and staining agents may alter the tissue's native properties, introducing artifacts that can affect the accuracy of histological analysis.

To address these challenges, label-free microscopy techniques [9, 10, 11, 12, 13] have garnered significant attention. These methods offer a non-invasive and rapid alternative to traditional staining procedures. By capturing label-free contrasts from unstained tissue samples, these microscopy techniques preserve the tissue's native state and biochemical integrity for subsequent analysis.

Moreover, virtual histology using label-free microscopy and deep learning algorithms [1, 8, 13, 14] has emerged as a promising solution to transform label-free contrasts into interpretable visualizations that resemble conventional histopathology stains. Deep learning models, such as Generative Adversarial Networks (GANs) [15], have demonstrated impressive capabilities in mapping label-free contrasts to corresponding stained images, enabling researchers to gain valuable insights without the need for chemical stains.

In this thesis, we focus on virtual histological staining of Photon Absorption Remote Sensing (PARS) images [16, 17], a label-free and all-optical modality that enables the simultaneous measurement of radiative and non-radiative relaxation of biomolecules. PARS offers a unique advantage by selectively imaging specific structures based on the absorption spectra of biomolecules like hemoglobin, DNA, and lipids without the need for extensive sample modifications.

For virtual staining, we employ and compare two different deep learning models: pix2pix [18] and cycleGAN [19]. Pix2pix is known for its accuracy in producing results; however, it requires pixel-level registration, which can lead to blurriness due to registration errors. To address this issue, we explored the cycleGAN model, which overcomes the registration problems of pix2pix. However, we observed that cycleGAN exhibits occasional occurrences of generating unrealistic and improbable outputs, resulting in features or patterns not present in the original training data, a phenomenon referred to as hallucinations. To mitigate this problem, we incorporated registered pairs of images to suppress hallucinations and enhance the overall performance of the model.

Despite these improvements, we noticed that both models tended to confuse various tissue structures, like cell nuclei and red blood cells. For a better understanding of the relationship between the optical and chemical properties of the tissue, we decided to incorporate additional information per pixel location. This involved utilizing features extracted from time-domain PARS signals. By adopting this approach, we achieved better labeling of tissue structures and, consequently, enhanced the colorization process. This step is crucial in achieving more accurate and realistic virtual histological stains.

Two different tissue types were involved in our research: human skin and mouse brain tissue sections. To begin, we focus on H&E staining, one of the most commonly used histological stains, to evaluate the performance of virtual staining with PARS microscopy. We aim to demonstrate the efficacy of our deep learning models in producing colorized images that resemble the results obtained through conventional staining methods. This comparison will provide valuable insights into the potential of virtual histology to revolutionize the field of histopathology.

Furthermore, beyond H&E staining, we explore the feasibility of translating PARS

data into Periodic Acid-Schiff (PAS) stain [20]. PAS staining is commonly used to highlight carbohydrates and glycogen in tissues. Successfully translating PARS data into PAS stain visualizations would expand the capabilities of virtual histological staining, offering researchers a versatile and powerful tool for studying a wide range of tissue types and biomolecules.

Through these endeavors, we demonstrate the potential of deep learning-based virtual histological staining with PARS microscopy for providing rapid and accurate insights into complex biological structures. The combination of label-free PARS microscopy and deep learning algorithms offers a promising approach to transforming the landscape of histopathology, opening up new avenues for medical diagnosis, research, and personalized healthcare.

The thesis is structured as follows. In Chapter 2, we provide the foundations of our research, including the background of PARS and virtual staining, along with a review of related virtual staining work. Chapter 3 presents the initial results obtained using PARS in combination with the pix2pix deep learning model.

Chapter 4 focuses on a comparative analysis of the performance between the pix2pix and cycleGAN models for virtual PARS histology. In Chapter 5, we elaborate on how we extended our work by intelligently extracting features from PARS time-resolved signals and incorporating them into the colorization process.

In Chapter 6, we present exploratory work, showcasing preliminary results of translating PARS data into PAS stain visualizations. Finally, Chapter 7 serves as a conclusion, summarizing our findings and suggesting potential directions for future research.

The principal thesis contributions can be summarized as follows:

- The thesis presents the initial publication of PARS virtual staining, where we employed the pix2pix deep learning model (Chapter 3).
- The utilization of multi-features for virtual staining of PARS images is a novel approach not previously explored in deep learning-based virtual histology (Chapter 5).
- The novel extension of our work to include PAS stain represents the first translation of PARS images to PAS stain (Chapter 6).

Additional contributions include:

- A comprehensive comparison of two well-established image-to-image translation models, evaluating their effectiveness in the virtual staining of the newly developed modality, PARS (Chapter 4).

- The extension of our deep learning models to accommodate more than three channels to accept time domain features for colorization (Chapter 5).
- The proposal of a robust framework for feature learning, extraction, selection, and color learning, tailored for PARS multi-feature fusion (Chapter 5).

# Chapter 2

# Preliminaries

## 2.1 Background

### 2.1.1 Histopathology

Histopathology is the fundamental microscopic examination tool in many domains, including cancer diagnosis and prognosis, surgical oncology, and drug discovery and development [6, 21, 22]. Tissue staining for microscopic imaging enabled histopathology by helping to distinguish the composition of tissue samples. The gold standard of histological imaging is bright-field microscopy of stained formalin-fixed paraffin-embedded (FFPE) thin tissue preparations, where the most common staining set is Hematoxylin and Eosin (H&E). Hematoxylin stains cell nuclei with deep blue-purple, while eosin stains cytoplasm and extracellular matrix with pink shades [23].

Developing these stained FFPE specimens for bright-field microscopic assessment relies on a laborious process of fixation, embedding, sectioning, and staining [7], a procedure that takes days to weeks to complete [8]. The fixation step involves treating the tissue sample with a fixative, typically formalin, to preserve its cellular structure and prevent decay. After fixation, the tissue is embedded in paraffin wax, which provides support and facilitates the sectioning process. Sectioning refers to the slicing of the embedded tissue into thin slices, typically around 5 to 10 micrometers thick, using a microtome. These thin sections are then mounted on glass slides and subjected to various staining techniques to enhance the visibility of cellular structures and specific molecules of interest.

While the FFPE method has long been considered the gold standard in histology, the lengthy process of sample preparation has led to the emergence of alternative techniques

such as frozen section (FS) histology [24]. FS allows for rapid intraoperative assessment, with turnaround times of approximately 20 minutes [25]. This technique involves freezing the tissue sample quickly and then cutting thin sections while it remains frozen. followed by the application of chemical staining, usually a modified H&E stains [25]. Although FS offers faster results, it presents certain limitations.

The sampling of specimens for FS is more challenging compared to FFPE, as it requires precise coordination between the surgeon and the pathologist. The rapid freezing process used in FS can introduce artifacts that adversely affect cellular morphology. These artifacts may distort the appearance of cells and tissues, making it more difficult for pathologists to accurately interpret the samples [25]. Consequently, the diagnostic reliability of FS may be compromised in certain cases.

Despite the advantages of FS in terms of speed, FFPE histochemical procedures remain the irreplaceable gold standard. The FFPE technique allows for a more comprehensive and detailed examination of tissue samples, providing pathologists with a clearer understanding of cellular morphology, architectural patterns, and the distribution of various biomarkers. This thorough analysis is crucial for accurate diagnosis, prognostication, and treatment planning in many medical disciplines. Therefore, despite the time-consuming nature of FFPE processing, it continues to be the preferred choice for routine histological assessments and research purposes.

Nonetheless, the current histology gold standard has several limitations that hinder its effectiveness. Histochemical staining not only delays sample assessment, but the sample preparation steps permanently alter the tissue composition with multiple chemical substances. As a result, tissues can only be stained once per section [23], requiring the preparation of independent slides for each desired staining contrast. This intensive and time-consuming process of preparing FFPE tissues can significantly delay the diagnosis of cancer and the assessment of margins, sometimes taking over a week. Consequently, the delays in obtaining these crucial histological results also impact the initiation of adjunct treatments and additional procedures that may be necessary for patient care. Given these limitations, there is a growing need for a clinically acceptable technique that can provide accurate, rapid, and reproducible histology-like images without the requirement for labeling or staining procedures. Such a technique would greatly enhance the efficiency and effectiveness of histological analysis in clinical settings.

### 2.1.2 Histological Staining Techniques

Histology staining techniques are used to enhance the visualization of specific components and structures within tissue samples under a microscope [26]. These techniques involve the application of various dyes, stains, or chemicals to the tissue sections, which selectively bind to different cellular components and produce distinct colors or contrasts. Examples of these stains include Hematoxylin and Eosin (H&E) and Periodic Acid-Schiff Stain (PAS), illustrated in Figure 2.1.

#### Hematoxylin and Eosin (H&E)

H&E is the most widely used staining technique in histology [27]. Hematoxylin stains the nuclei of cells blue-purple, while eosin stains the cytoplasm and extracellular components pink [23], as shown in Figure 2.1 (b). H&E staining provides general information about tissue architecture and cellular morphology.

Hematoxylin is a natural dye derived from the heartwood of the logwood tree [28]. It acts as a basic or cationic dye, meaning it binds to negatively charged components in tissues. Hematoxylin stains the nuclei of cells as well as other acidic structures, such as nucleoli and chromatin, in shades of blue or purple [28]. The intensity of the stain can be adjusted by the duration of hematoxylin exposure [29].

Eosin is an acidic or anionic dye that binds to positively charged components in tissues [29]. It stains the cytoplasm, collagen fibers, and extracellular matrix in shades of pink or red [23]. Eosin staining provides contrast to the blue-stained nuclei, allowing for the differentiation of different cell types and tissue structures.

Tissue examination using H&E staining facilitates the visualization of cellular details, encompassing characteristics such as cell shape, size, and arrangement [30]. This staining technique offers valuable insights into the organization of tissues, enabling the identification of glandular structures, blood vessels, and epithelial layers [27]. H&E staining has been widely used for the diagnostic purposes of numerous diseases, including cancer [31].

In the realm of histopathological research, H&E staining holds significant importance. Researchers extensively analyze H&E-stained tissue sections to investigate diverse aspects of tissue development, disease mechanisms, treatment responses, and other biological processes [32].

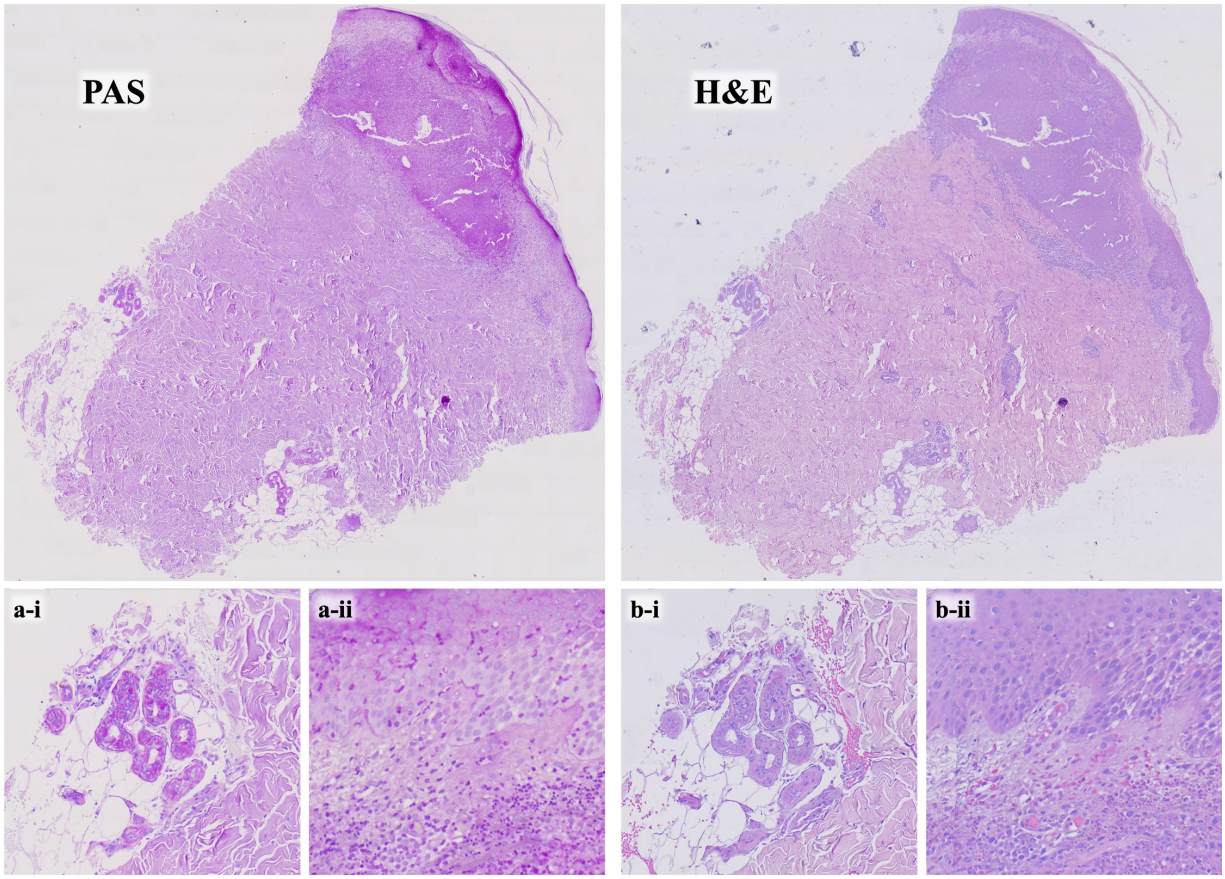



Figure 2.1: Images of H&E and PAS histochemical stains of the same human skin fungus tissue section. (a) PAS, stains fungus with magenta, cytoplasm and extracellular structures with shades of pink, and nuclei with blue-violet. (b) H&E, stains cytoplasm and extracellular structures with shades of pink and nuclei with blue-violet. (i) Part of the tissue showing sweat glands and connective tissue. (ii) Part of the dermis layer encompassing cell nuclei and fungus.

#### Periodic Acid-Schiff Stain (PAS)

PAS is a widely utilized histological staining technique in pathology and medical research. It is employed to detect carbohydrates, glycogen, mucosubstances, and glycoproteins in tissue sections [20]. This staining method involves the reaction between periodic acid and carbohydrates, resulting in the formation of aldehydes that react with Schiff reagent to produce a colored product visible under a microscope [33]. Tissues stained with PAS exhibit blue coloration of the nuclei, while glycogen and fungi exhibit a magenta hue [34], as illustrated in Figure 2.1 (a). PAS staining plays a critical role in the diagnosis of various diseases, including glycogen storage diseases [35] and fungal infections [36].

The procedure for PAS staining involves deparaffinization of tissue sections, oxidation using periodic acid to produce aldehydes, application of Schiff reagent to react with the aldehydes, washing to eliminate excess reagents, optional counterstaining with hematoxylin to enhance contrast, and finally, dehydrating and mounting the samples for microscopic examination. [37].

### 2.1.3 Label-Free Microscopy

Label-free microscopy techniques have garnered significant attention in the field [9, 10, 11, 12, 13], as they enable the acquisition of informative contrast from unstained samples while maintaining the structural and biochemical integrity of the specimens for subsequent analysis. These label-free contrasts can be extracted and transformed into familiar formats, resembling conventional histopathology stains, through a process known as virtual staining. This approach allows for the translation of label-free contrasts obtained from different optical microscopes into interpretable images using deep learning algorithms. Various microscopy techniques, such as quantitative phase imaging (QPI) [38], reflectance confocal microscopy [39], photoacoustic microscopy [40], and autofluorescence microscopy [41], have been coupled with deep learning algorithms [8, 9, 13, 42] to intelligently present the rich information in meaningful ways.

Each type of microscopy has a varying degree of specificity toward various biomolecules. As a microscope becomes more specific to the target biomolecules, the reliance on deep-learning models for inference decreases. For example, autofluorescence microscopy shows promise in capturing extranuclear elements, but it may not provide a direct measurement of nuclear contrast, as cell nuclei do not exhibit measurable autofluorescence [43]. This limitation can impact the accuracy of deep learning-based estimation of the coloring of nuclear structures.

Autofluorescence is a form of radiative relaxation and re-emittance of lower energy photons following the absorption of light. However, a portion of the absorbed energy can also undergo relaxation via temperature and pressure [16, 17, 40]. These are termed non-radiative relaxation pathways, and they provide additional contrast and hold significant value in capturing critical structures, including cell nuclei (DNA) [17].

One newly developed modality, Photon Absorption Remote Sensing (PARS) microscopy, previously known as Photoacoustic Remote Sensing [16, 17], can simultaneously measure the radiative (e.g., autofluorescence) and non-radiative (temperature and pressure) relaxation of biomolecules. By leveraging the unique absorption spectra of biomolecules like hemoglobin, DNA, and lipids, PARS enables direct selective imaging without the need for extensive systematic modifications [44, 45, 46]. PARS has emerged as a powerful method that utilizes a non-contact, all-optical architecture to visualize the intrinsic absorption contrast of biological tissue structures.

### 2.1.4 PARS

PARS operates based on the principles of photoacoustic imaging, which utilizes laser-induced acoustic waves to detect and map the distribution of light-absorbing molecules within biological tissues. The development of first-generation of PARS builds upon the advancements of previous photoacoustic techniques. Furthermore, the second-generation PARS has significantly improved its imaging capabilities through enhancements in hardware, signal processing, and data acquisition methods. It introduced a novel approach that combines radiative and non-radiative contrasts to provide insights into total chromophore absorption, leading to the adoption of the term "Photon Absorption Remote Sensing." This unique imaging method offers a comprehensive understanding of optical properties, differentiating itself from other techniques, as far as our knowledge extends.

These advancements have enabled more accurate and detailed visualization of tissue structures, providing valuable insights into the composition of biological samples. In the following sections, we will explore the fundamentals of photoacoustic techniques and their applications in imaging and delve into the developments of the second-generation PARS.

#### History and Physical Mechanisms of Photoacoustics (PA)

The photoacoustic (PA) effect (also called the optoacoustic effect) was discovered by Alexander Graham Bell in 1880 [47] during his experiments with transmitting sound at

long distances. Bell's well-known invention, the photophone, resulted in an important observation: the ability of electromagnetic waves (including visible and non-visible light) to produce sound waves when applied to materials. Consequently, he developed a new device called the "spectrophone" to utilize this effect for spectrally identifying different materials.

The physics behind this phenomenon [40, 47] can be explained as follows. When light is absorbed by a sample material, it heats up and expands (thermoelastic expansion). Accordingly, the pressure around the sample changes and propagates as acoustic waves. A varying light source, such as modulated or pulsed light, is critical to obtaining the PA effect. The generated wide-band acoustic waves are measured using adequate detectors such as microphones or piezoelectric sensors.

### Photoacoustic Imaging (PAI)

Recently, the photoacoustic effect has been utilized in biomedical imaging, namely, photoacoustic imaging (PAI) or photoacoustic tomography (PAT) [48, 49]. PAI has been growing rapidly in the field of biomedical imaging research; it has become a very useful tool to perform correlative multi-scale studies in the biomedical field [49] for its capability to bridge the microscopic and macroscopic imaging domains. PAI is a hybrid modality that benefits from both sound and light properties. In comparison to other all-optical imaging techniques, PAI makes use of the fact that sound waves attenuate less than light does within tissues, allowing the capture of deeper images with high spatial resolution [50]. On the other hand, ultrasonic imaging techniques [51, 52, 53] can also achieve high spatial resolutions at depths, though the technique is unable to obtain the same level of contrast as optical modalities that, for instance, can effectively image hemoglobin oxygen saturation [54].

PAI offers several benefits, including the ability to produce high-contrast and high-resolution images without the use of ionizing radiation [55]. Moreover, according to the system configuration and imaging depth, PAI can achieve a spatial resolution ranging from nanometer to millimeter for imaging structures from organelles to organs and even small animal bodies and humans [49]. Different PAI modalities can cover imaging depths from 100 $\mu m$ up to 7 cm [49]. Besides, it can acquire signals and form images with near-real-time performance [56]. PAI can capture almost all structures that have adequate optical absorption profiles at the chosen wavelength [48]. Endogenous and exogenous contrast agents [57]were utilized to image such structures. They can boost imaging quality and provide complementary structural and functional information about various biological tissues.

Endogenous chromophores have helped PAI perform label-free imaging and reveal

anatomical, functional, and histological properties of tissues [57]. Examples of absorbing endogenous contrast agents include DNA/RNA, hemoglobin, myoglobin, cytochromes, lipid, melanin, and water. The common use of DNA and RNA absorbers is the imaging of cell nuclei at ultraviolet wavelengths [58], while hemoglobin helps in vascular imaging at wavelengths around the visible and near-infrared spectra [59]. Likewise, melanin is used for melanoma tumor imaging in the near-infrared spectral region [60]. Differentiating between one endogenous agent and the other is achievable with spectral measurement (if the optical fluence $F$ is known). On the other hand, the use of exogenous contrast agents is planned to optimize their light absorption to match specific wavelengths; this has helped PAI achieve molecular and cellular specificity [57]. They are preferred to endogenous agents when higher detection sensitivity is required or when targeting specific molecules, such as antibodies. Exogenous contrast agents include organic dyes, fluorescent and non-fluorescent proteins, and nanoparticles [61].

General examples of PAI utilization in clinical and preclinical applications include brain vasculature and functional imaging [62], visualization of blood oxygen saturation [54], histology-like tissue imaging [63], detection of skin cancer [64], gene expression [61], and eye imaging with a multi-modal system [65,66,67]. It has also been utilized for non-medical purposes such as defect detection of metal surfaces [68] and exploring unrevealed features of paintings [69]. Furthermore, exogenous contrast agents can be used instead, and the laser is tuned accordingly to obtain highly detailed images and enhance the visualization of certain structures like blood vessels [70, 71]. More specific applications for each PAI modality are presented later in this section.

### Photoacoustic Imaging Principles

The general PAI mechanism [72] can be described as follows. The tissue of interest is illuminated using a short laser pulse, where different chromophores in the tissue absorb the light energy according to their absorption coefficient ($\mu_a$) at the chosen excitation wavelength, and then they start to vibrate, generating acoustic waves. These waves are then captured and passed to a software program to reconstruct 2-D or 3-D images from the detected photoacoustic signals, giving anatomical, functional, and pathological information about the tissue. PAI can take advantage of endogenous chromophores such as oxygenated and deoxygenated blood, melanin, water, and fats to provide label-free imaging. Figure 2.2 depicts the absorption coefficient ($\mu_a$) spectrum of major chromophores in biological tissues [3].

PAI's working principle [72] relies on the photoacoustic effect, presented earlier in this section. A short-pulsed light source is used to illuminate the target structure. A portion

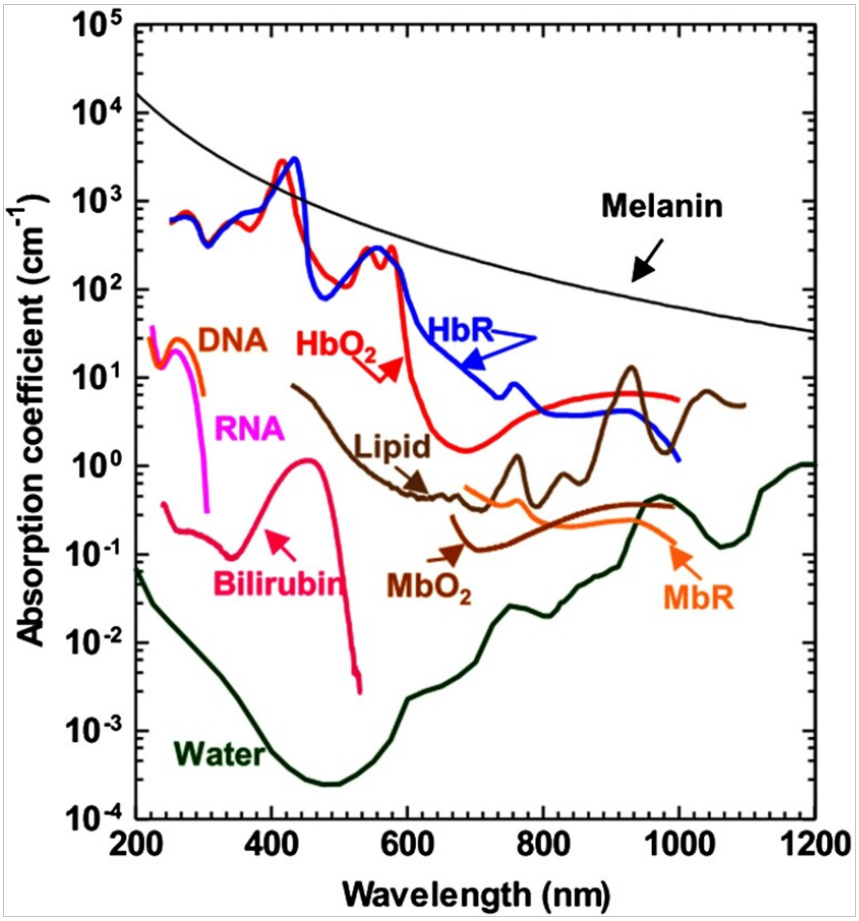



Figure 2.2: Absorption spectra of the major endogenous chromophores of biological tissue. Source: Reprinted from [3].

of this light is absorbed by the sample and partially converted into heat, leading to a rise in temperature. This increase in temperature is further translated into a pressure rise, which is propagated as an ultrasonic wave in the order of MHz. This wave is referred to as a photoacoustic wave. In PAI, the PA signal is detected using an ultrasonic transducer composed of either a single element or an array of elements.

Two important time scales exist during laser excitation: thermal ($\tau_{th}$) and stress ($\tau_s$) relaxation times. When ($\tau_{th}$) and ($\tau_s$) are larger than the laser pulse duration, the excitation is said to be in thermal and stress confinement, respectively, and thus the heat conduction and stress propagation are negligible during a laser pulse [72]. Consequently, the fractional volume expansion can be neglected, and the local initial pressure rise $p_0(x)$ at location $x$ immediately after the laser pulse is given by [72]

$$p_0(x) = \frac{\beta T}{k}, \tag{2.1}$$

where $\beta$ denotes the thermal coefficient of volume expansion ($K^{-1}$), $T$ is the temperature rise (K), and $k$ represents the isothermal compressibility ($Pa^1$). The temperature rise $T$ is given by [72]

$$T = \frac{\eta_{th} A_e}{\rho C_v}, \tag{2.2}$$

where $\eta_{th}$ is the absorbed light ratio transformed into heat, $A_e$ is the specific optical energy deposition ($J/m^3$) and $C_v$ is the specific heat capacity at constant volume. The Gruneisen parameter [72] is defined as

$$\Gamma = \frac{\beta}{k\rho C_v} = \frac{\beta v_s^2}{C_p}, \tag{2.3}$$

where $v_s$ denotes the velocity of sound in the medium. And since $A_e$ corresponds to the local optical fluence $F$ for a single-photon optical absorption (measured in $J/cm^2$), therefore the local pressure rise immediately after the excitation laser pulse is given by [72]

$$p_0(x) = \rho \eta_{th} \mu_a F, \tag{2.4}$$

where $\mu_a$ is the optical absorption coefficient measured in $cm^{-1}$. Since $\Gamma$ and $\eta_{th}$ are

commonly assumed to be constants [73], Eq. 2.4 implies that the initial pressure is proportional to $\mu_a$ and $F$, where $\mu_a$ can be computed as $F$ is known and $p_0(x)$ is measurable. The creation of the initial pressure $p_0$ results in acoustic waves propagating at the speed of sound within the sample being imaged. The mathematical formulation for the generated time-domain signal can be found in [72].

### Photoacoustic Imaging Modalities

PAI is categorized into two main imaging systems: photoacoustic computed tomography (PACT) and photoacoustic microscopy (PAM) [48,49,74]. In this section, we briefly discuss the key differences between the two modalities, present some recent applications, and introduce relevant technical specifications.

The main difference between the two modalities is the technique by which the image is formed [49, 74], which in turn leads to significant differences in their system design. To capture the initial pressure distribution $p_0(x, t)$ at location $x$ and time $t$ caused by the absorption of light in the tissue, PACT systems use reconstruction-based algorithms, while PAM systems rely on focused scanning image formation. PACT uses widefield light illumination, and the acoustic signal is collected at multiple locations using an ultrasonic transducer array, whereas PAM is based on focused excitation light beam scanning and a single focused detector. A high pulse repetition rate for rapid imaging is key in PAM systems, while PACT needs high pulse energy [49]. PAM does not need high energy ($<1$ $\mu$ J for optical-resolution PAM and $<1$ mJ for acoustic-resolution PAM) as the excited region is relatively small, contrary to PACT ($>10$ mJ), which optically excites wider regions. In general, PACT can penetrate better (used in diffuse regime imaging) and image faster, but at the cost of lower spatial resolution when compared with state-of-the-art PAM techniques. Since the proposed work is concerned with studying biological tissues at a microscopic level, PAM and its variations are our focus in this thesis. More details on PAM variants and their technical specifications are introduced below.

PAM is an important type of PAI that performs raster scanning to acquire depth-resolved signals and then directly recovers the images from them. An A-scan in PAM is the direct depth-resolved acoustic signals (1D), a B-scan is acquired by scanning along the x-axis while fixing the y-axis position (2D), and a C-scan is achieved by raster scanning a grid of A-scans (3D). PAM systems take advantage of the better scattering performance (two to three orders of magnitude) of the acoustic waves compared to that of the light waves within the sample being imaged, and thus, PAM overcomes the optical diffusion limit [72]. In comparison with other pure optical modalities, such as optical coherence tomography and confocal microscopy, that provide scattering contrast, PAM provides rich

optical absorption contrast imaging. Consequently, PAM can perform molecular imaging with the aid of either endogenous (label-free) or exogenous contrast agents, where it can image almost all the molecules rather than a small fraction of them when compared to fluorescence-based techniques (like multi-photon microscopy). An exceptional advantage of PAM is its capability to concurrently image in vivo functional, structural, and flow dynamic contrasts.

The two major PAM implementations [56,75,76] are optical-resolution PAM (OR-PAM) and acoustic-resolution PAM (AR-PAM). The main difference between both modalities is that OR-PAM takes advantage of using a tighter optical focus, while AR-PAM adopts the operation of a tighter acoustic focus. Another difference is in the way each of the techniques is implemented: OR-PAM can be operated in any of the three confocal microscopy modes (reflection, transmission, and double-illumination modes), while AR-PAM preferably operates in reflection mode only because its primary usage is for thick structure imaging [76]. Figure 2.3 shows the setup of the main PAI implementations: OR-PAM, AR-PAM, and PACT [77]. As discussed later in this section, our work scope is related to OR-PAM; hence, we introduce below its main features and recent advances.

**Lateral Resolution.** OR-PAM outperforms AR-PAM in spatial resolution because the optical focus can be adjusted to be much tighter than the acoustic one due to the smaller optical wavelengths. Nevertheless, AR-PAM is better when dealing with deep structures because of the acoustic wave's ability to penetrate more deeply into the tissue than light does.

For an OR-PAM operating in reflection mode, both the laser illuminating the tissue and the detecting wide-band ultrasonic transducer are co-aligned. Since the optical spot is about 10 times tighter than the acoustic focus spot, the detection sensitivity across the spot is almost the same.

The lateral resolution for OR-PAM can be defined as the optical focus spot size at the diffraction limit and is expressed as [56]

$$R_{L,OR} = 0.51\frac{\lambda_0}{NA_0}, \tag{2.5}$$

where the constant 0.51 represents the full width at half maximum (FWHM) of the light intensity focal spot, $\lambda_0$ denotes the optical wavelength, and $NA_0$ is the numerical aperture of the optical objective. From Eq. 2.5, the lateral resolution can be improved by either increasing the numerical aperture $NA_0$ or altering the excitation wavelength to be shorter.

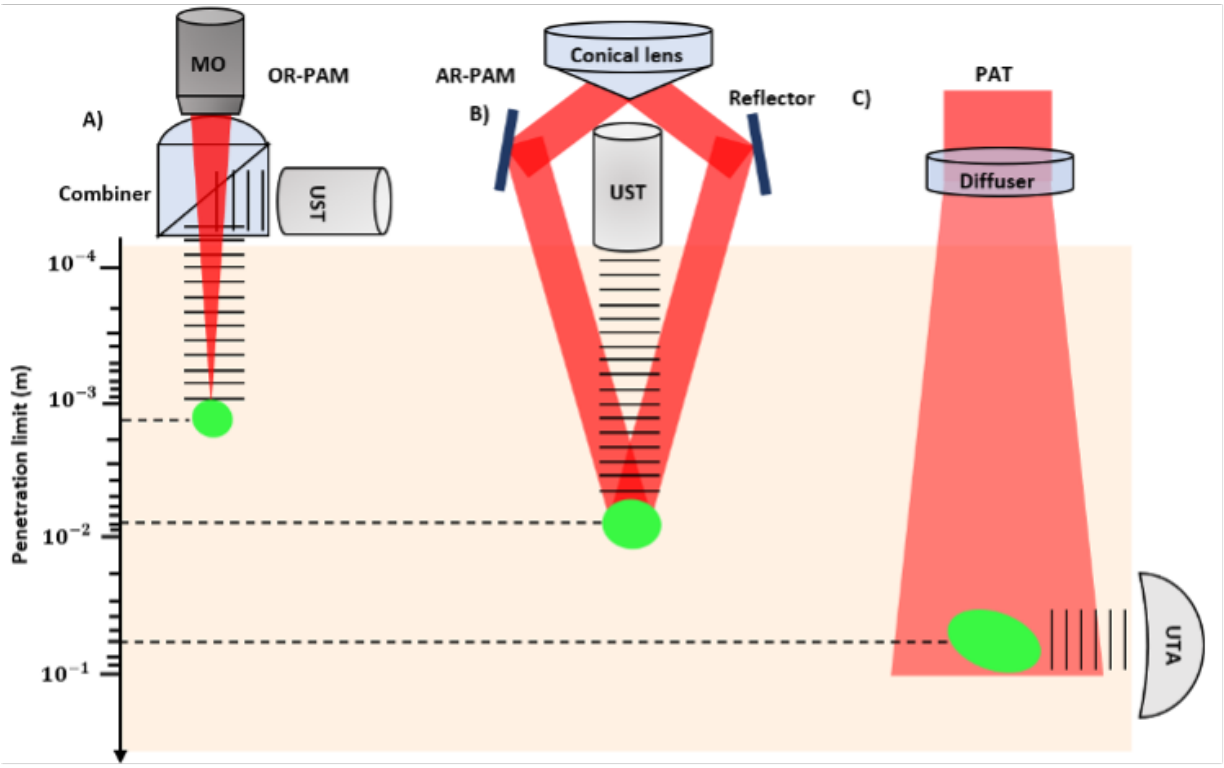



Figure 2.3: Different implementations of PAI and their penetration limit (A) Reflection-mode OR-PAM system. (B) AR-PAM system. (C) PACT system with widefield illumination. UST: ultrasonic transducer. UTA: ultrasonic transducer array. Source: Reprinted from [77].

It should be considered that these changes, on the other hand, will affect the maximum imaging depth.

**Axial Resolution.** Both OR-PAM and AR-PAM axial resolutions can be computed by [56]

$$R_{A,OR/AR} = 0.88v/\Delta f_c \tag{2.6}$$

where $\Delta f_c$ is the bandwidth of the PA signal (proportional to the central frequency $f_c$) and $v$ is the sound speed in the medium. Again, because lower frequencies travel deeper, the PA signal frequency range decreases with depth, resulting in lower axial resolution. OR-PAM can reach higher axial resolution [56] with sufficiently wide bandwidth transducers. OR-PAM is preferred for imaging in the quasi-ballistic regime as it can resolve organelles and single cells, while AR-PAM is preferred in a quasi-diffusive regime.

**Imaging Depth.** OR-PAM's maximum imaging depth is on the order of 1 mm in soft tissue, which is about one transport mean free path [78], due to optical scattering. The optical attenuation inside the tissue includes scattering and absorption, where scattering does not strongly rely on wavelength. As for AR-PAM, the main limitation is the ultrasonic attenuation, which depends on the acoustic frequency. In addition to scattering, the imaging depth is bounded by light absorption which highly depends on the laser beam wavelength.

**Imaging Speed.** In addition to the need to capture and process images as fast as possible, imaging speed is also important for avoiding motion artifacts (like breathing) while imaging details and targeted motions (like blood flow). A typical OR-PAM system with a mechanical scanning approach captures a B-scan at a rate of 1 Hz [76]. This means that imaging an area of 1 x 1 $mm^2$ takes about 7 min using OR-PAM and 1 min with dark-field AR-PAM. Since mechanical scanning is considered a slow method, one way to enhance the scanning speed is to use 2D galvanometer mirrors instead. A B-scan rate of 400 Hz was then achieved using MEMS-OR-PAM while scanning a 3 mm range. The addition of MEMS to the system helps maintain high sensitivity while simultaneously achieving a wide field of view [79].

PAI modalities have been adopted in a wide range of applications. OR-PAM can resolve single cells at sub-micrometer resolution and has thus been successfully used in single-cell functional and structural imaging [80]. Also, imaging of the DNA and RNA [58], which

are present in the cell nuclei, and cytochromes within the cytoplasm, is achieved without the need for staining. This is important for in vivo histologic imaging. It can also perform blood cell morphology imaging [81] to detect abnormalities and differentiate between cell types according to the frequency information extracted from the PA signal. OR-PAM can also image single melanosomes in melanoma cells [82]. As for AR-PAM, it has been successful in studying melanoma progress in small animal brains at high depths [83]. On the other hand, PACT systems are usually used for imaging an organ, like a whole mouse brain [84] or the whole body of a small animal [85].

PAI commonly uses a nanosecond pulsed laser [61], and the wavelength is chosen to be anywhere from ultraviolet to near-infrared depending on the chromophores of interest. As for the detection side, several types of detectors have been used, also depending on the intended application. Some approaches use contact-based detection mechanisms such as piezoelectric transducer, capacitive micromachined ultrasound transducer (CMUT), micro ring resonator (MRR), and Fabry-Perot interferometer (FPI) [86]. Other approaches are non-contact-based coupling techniques like air/gas-coupled detection methods [87]. Later, a more advanced technique based on remote detection, photoacoustic remote sensing (PARS) [16], was introduced.

Compared to other microscopy techniques, PAI or, to be more specific, PAM, is not considered an all-optical imaging technique because of its need to be in physical contact for detection with the sample through a coupling medium [88]. This limits its flexibility to be combined with other pure optical imaging microscopes. PAM has been known for its high sensitivity, resolution, and imaging contrast; thus, its utilization to further provide complementary imaging contrast when combined with other modalities has recently attracted a lot of research groups [65, 77, 89]. Another problem of contact-based methods is the degree of loss resulting from poorly matched interfaces (such as air and tissue). This is an issue because it forces portions of the generated pressure wave to be reflected away from the transducer and thus reducing the detection sensitivity. Furthermore, some applications, such as ophthalmology, are negatively affected by contact-based methods in several ways [77]. Firstly, a downside is the patient discomfort and the risk of infection, and secondly, there is the effect of the eye movements on the coupling efficiency and therefore the image quality. Moreover, the opaque nature of ultrasonic transducers prevents them from being practical enough for ocular imaging, in addition to other applications such as wound assessment and endoscopic operations [88]. Furthermore, in brain imaging, a special orientation of the transducer is preferred for the solution stability and maximum scanning efficiency, which is not convenient for patients while on a surgical bed.

For all these reasons, non-contact detection methods paved the way for utilizing the PA effect in more clinical applications [56], either on its own or combined with other modalities.

To achieve the aforementioned goals, air- and gas-coupled transducers have been developed. They have been widely deployed in non-medical areas such as material characterization and quality control in the food industry, among others [87]. It can be deduced that the main problem of such methods is the challenge of matching air and solid impedances to avoid ultrasonic signal attenuation [56]. Recently, non-contact optical detection methods have started to replace other techniques because of their significant performance in accurately capturing ultrasound signals [90,91]. All-optical detection methods can be categorized into two main groups: interferometric and non-interferometric detection methods [56].

Although non-contact interferometric detectors have overcome most of the challenges introduced by contact-based methods, they have also created other technical problems [92]. Interferometric detection's common issue is the reduced image quality, as it mainly relies on acquiring the vibrations occurring on the tissue surface. Besides, the focus of the excitation and detection beams needs extensive effort to be adjusted, limiting its application for in vivo imaging. Consequently, the need for a more sophisticated detection method is currently in great demand.

Recently, a non-contact, non-interferometric approach for photoacoustic signal detection, namely, photoacoustic remote sensing (PARS) [16, 44, 45, 46, 93, 94] has been introduced by HajiReza et al. in 2017. PARS targets the construction of high-resolution and high-contrast images while avoiding most of the complications caused by contact-based and interference-based methods. The working principle with regards to excitation is similar to OR-PAM but the detection mechanism is different, as described below. In PARS microscopy, a nanosecond laser beam is used to excite the tissue. The chromophores then absorb the optical energy, and their temperatures rise, leading to a micro-scale initial pressure $p_0(x)$ at location $x$ in the order of megapascal through thermo-elastic expansion. This very high magnitude of the initial pressure results in a refractive index $n_0(x)$ mismatch (elasto-optic modulation) within the absorber, and therefore changes the back-reflected detection (also called a probe or interrogation) beam intensity. The amplitude of the reflected signal is proportional to the amount of light absorbed by the chromophores at the excitation wavelength. To capture such optical signals, the continuous-wave low-coherent detection beam is co-focused and co-scanned with the excitation beam spot. The confocal detection configuration helps in harvesting the PA signal at the subsurface origin, where maximum pressure takes place. Unlike the interferometric methods discussed earlier in this section, PARS focuses on the changes happening to the back-reflected signals while rejecting any phase information that might have occurred. An illustration of a typical PARS setup is shown in Figure 2.4 [16].

The local initial pressure can be computed with the assumption that the thermal confinement (introduced earlier in this section) and the optical focusing conditions are ap-

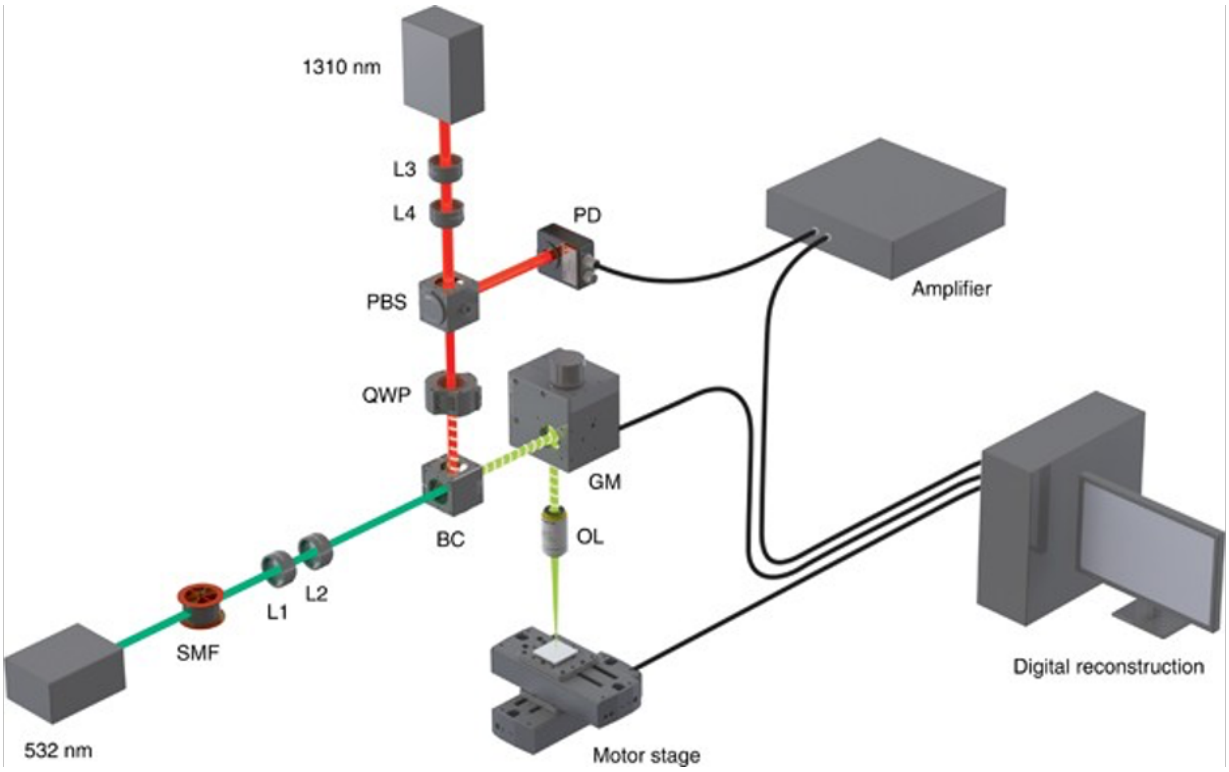



Figure 2.4: The typical setup of PARS system. L: lens, PBS: polarized beam splitter, PD: photodiode, QWP: quarter-waveplate, BC: beam combiner, SMF: single-mode fiber, GM: galvo-scanner mirrors, OL: objective lens. Source: Reprinted from [16].

plied [16]. This change in pressure is a step shape wave and it modulates the original medium refractive index $n_0(x)$ by an amount $\delta_n(x)$, the elasto-optical relation [95] is given by [16]

$$n(x,t) = n_0(x) + \delta_n(x,t) = n_0(x)\left(\frac{1+\eta n_0(x)^2 p(x,t)}{2\rho v_a^2}\right), \tag{2.7}$$

where, $\eta$ denotes the elasto-optic coefficient (about 0.32 for water), $p(x,t)$ represents the pressure field, $\rho$ is mass density and $v_a$ is the sound speed in a medium.

The perturbed intensity reflection coefficient as a result of the initial pressure at the absorbing media boundaries immediately after an exciting pulse can be defined as [16]

$$\Delta R = \delta_n\left(\frac{2Re(n_1-n_2)}{\mid n_1+n_2\mid^2} - \frac{2Re(n_1+n_2)}{\mid n_1+n_2\mid^2}\left|\frac{n_1-n_2}{n_1+n_2}\right|^2\right) + \mathcal{O}(\delta_n), \tag{2.8}$$

where $\delta_n$ is mostly assumed to be real, $\left|\frac{n_1-n_2}{n_1+n_2}\right|^2$ represents the unperturbed intensity-reflection coefficient, $n_1$ is the absorbing medium refractive index, $n_2$ is the non-absorbing medium refractive index, and $\mathcal{O}(\delta_n)$ denotes the terms that depend on $\delta_n$ higher orders. It can be shown that $\Delta R$ depends on higher orders of $\delta_n$ when both refractive indices are approximately the same; otherwise, its modulation is proportional to $\delta_n$ and thus proportional to the initial pressure.

PARS has been able to achieve a lateral resolution of 2.7 $\mu m$ for in vivo mouse ear images of a large field of view, with excitation beam surface fluence $\sim 12$ times lower than the ANSI limits and detection beam surface average power $\sim 8$ times lower than the ANSI limits [16]. Furthermore, PARS has achieved non-contact deep imaging ($> 1.2$ $mm$ in vivo) beyond the excitation wavelength's transport mean-free path [96]. It has also managed to obtain real-time functional imaging, such as measuring oxygen saturation in blood flow phantoms and in vivo mouse ear microvasculature [93]. Since the typical PARS system cannot produce depth-resolved information in a single acquisition, optical sectioning is needed to obtain such information.

In 2022, the development of the second-generation PARS [17] was motivated by the understanding that energy absorbed by a sample can undergo relaxation through temperature and pressure processes (non-radiative relaxation), in addition to radiative emissions like autofluorescence. These different contrasts are valuable for capturing various important structures, including extracellular structures, red blood cells, and cell nuclei (DNA) [17].

By combining both non-radiative and radiative contrasts, second-generation PARS enables improved capture and visualization of these structures, providing a comprehensive understanding of the sample's optical properties. Hence, the new terminology "Photon Absorption Remote Sensing" is introduced.

Second-generation PARS microscopy allows for the simultaneous capture of all three contrasts in a single excitation event in a label-free and non-contact manner. To the best of our knowledge, no other single imaging technique currently offers this unique capability [17]. PARS achieves this by using a pulsed excitation laser to induce both radiative and non-radiative changes in the sample, leading to modifications in the local optical properties. These changes are detected by a secondary probe beam, which measures variations in backscattering intensity caused by non-radiative absorption. Additionally, the intensity of the detection laser's pre-excitation scattering is utilized to collect optical scattering contrast.

### 2.1.5 Virtual Staining

Virtual staining [1, 8, 13, 14] is a process that involves transforming label-free optical microscopy data into visualizations that resemble conventional histopathology stains, such as H&E. It is achieved through the use of algorithms, particularly deep learning algorithms [1, 13, 97], that simulate the effects of chemical staining on unlabeled tissue samples.

Label-free optical microscopy techniques capture intrinsic contrasts from unstained tissue samples without the need for exogenous dyes or stains. These techniques provide valuable information about the structural and biochemical properties of the tissue. However, the interpretation of label-free images by pathologists, who are trained to analyze histochemically stained samples, can be challenging.

Virtual staining bridges this gap by generating H&E-like visualizations from label-free microscopy data. Deep learning algorithms are trained on a large dataset of paired [1, 13] or unpaired [97] label-free images, $X$, and corresponding chemically stained images, $Y$. These algorithms learn the mapping, $G : X \rightarrow Y$ between the label-free contrasts and the corresponding staining effects, where $G$ is the trained image generator, as discussed in the following subsections. Once trained, the algorithms can apply this mapping to new label-free images, generating virtual stains that resemble traditional histopathology stains.

The goal of virtual staining is to facilitate the interpretation of label-free images by providing pathologists with familiar visual representations. This approach allows for the integration of label-free imaging techniques into routine clinical practice, improving the

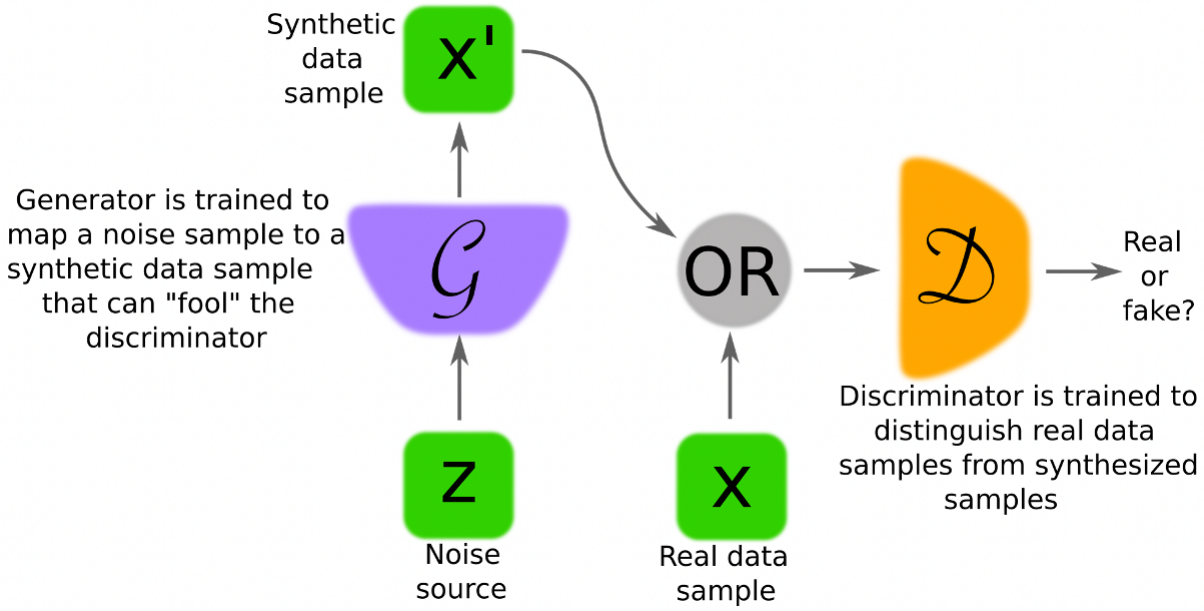



Figure 2.5: Overview of a GAN training algorithm. Source: Reprinted from [4].

diagnostic utility of small tissue specimens and reducing the variability associated with traditional staining processes.

### 2.1.6 Generative Adversarial Networks (GANs)

In this work, we choose to adopt a general deep learning-based image-to-image translation method to perform virtual staining using PARS images. The method is based on a generative adversarial network (GAN) [15]. GAN is a type of deep learning generative model that consists of two main components: a generator and a discriminator. GANs are specifically designed for generating synthetic data that resembles a given training dataset. The generator generates fake data samples, while the discriminator's role is to distinguish between the generated (fake) samples and the real samples from the training dataset. An illustration of the training algorithm is shown in Figure 2.5.

The key idea behind GANs is that the generator and discriminator are trained in an adversarial manner. The generator aimed to produce realistic samples that could fool the discriminator, while the discriminator learns to become increasingly accurate in distinguishing between real and fake samples. This adversarial training process drives the

generator to improve its ability to generate increasingly realistic samples.

GANs have gained popularity in the field of deep learning due to their ability to generate high-quality synthetic data across various domains, such as inpainting [98], super-resolution [99], and colorization [100, 101, 102]. They are particularly effective at capturing the complex distributions and structures present in the training data.

In a standard GAN architecture, the generator takes random noise as input and generates synthetic data [15]. However, in certain applications, it can be beneficial to condition the GAN on input data. This is known as a conditional GAN (cGAN) [103]. In a cGAN, both the generator and discriminator are conditioned on the additional input information, such as class labels or specific attributes, to guide the generation process.

Conditioning the GAN on input data provides greater control over the generated samples. In the case of virtual staining, this can involve conditioning a cGAN on label-free microscopy data, resulting in the generation of virtual stains tailored to the specific characteristics of the input data.

The training process involves using paired [1, 13] or unpaired [97, 104, 105] datasets of label-free and stained images to train the GAN. Through this training, the generator learns to generate virtual stains that mimic the staining effects observed in real histopathology, conditioned on the given input information. By leveraging the power of GANs, virtual staining offers a solution to the challenge of converting label-free microscopy data into visually interpretable images that resemble traditional histopathology stains.

In our work, we utilize two GAN-based models to address image-to-image translation tasks. The first model is called pix2pix [18], which operates on paired images. The second model is cycleGAN [19], which can handle unpaired image datasets. Chapter 4 of the thesis includes a comparison of the two models. In the subsequent subsections, we will introduce the two models and explain their functions.

### Pix2pix

Pix2pix [18] focuses on paired image datasets, where each input image is associated with a corresponding target output image, enabling precise translation. It learns the mapping from an input $x$ and a random noise vector $z$ to an output $y$. By utilizing a conditional GAN framework, both the generator $G$ and discriminator $D$ are conditioned on input images. This allows pix2pix to generate output images that are semantically similar to the corresponding input images.

The objective function of pix2pix consists of two main components: the adversarial loss (GAN loss) and the pixel-wise loss [18]. The adversarial loss is calculated using the

standard cGAN objective, which minimizes the cross-entropy loss between the predicted probability of the discriminator for real images and the true labels, as well as the predicted probability for fake images and false labels. The cGAN adversarial loss can be expressed as [18]

$$\begin{aligned}\mathcal{L}_{cGAN}(G, D) =& \mathbb{E}_{x,y}[\log D(x, y)] + \\ & \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z))].\end{aligned} \tag{2.9}$$

The pixel-wise loss measures the pixel-level similarity between the generated output images and the target images. It is computed as the mean absolute difference or mean squared difference between corresponding pixels in the generated and target images. The pixel-wise loss encourages the generator to produce visually similar output images to the target images at a finer level of detail. Here, we choose to follow [18] and use L1 loss as pixel-wise loss,

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}\left[\|y - G(x, z)\|_1\right]. \tag{2.10}$$

With these combined losses, pix2pix generates visually appealing and realistic output images faithful to the input. The overall objective is given by [18]

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \tag{2.11}$$

where $\lambda$ is a hyperparameter that controls the trade-off between the adversarial and pixel-wise losses.

### CycleGAN

In contrast, cycleGAN [19] is designed to handle unpaired image datasets, where input and output images are not explicitly paired. Instead of focusing on precise correspondence, cycleGAN aimed to learn the underlying mapping between two domains. It achieves this through the introduction of cycle consistency. The objective function of cycleGAN comprises three main components: the adversarial loss, the cycle consistency loss, and optionally, the identity loss.

Similar to pix2pix, the adversarial loss in cycleGAN aimed to make the generated (translated) output images from one domain ($X$) appear as realistic as possible to the

discriminator. The discriminator's objective is to distinguish between real images from the target domain ($Y$) and translated images.

The objective [19] is to learn mapping functions between two domains, $X$ and $Y$, using training samples $\{x_i\}_{i=1}^{N}$ from domain $X$ and $\{y_j\}_{j=1}^{M}$ from domain $Y$. The data distribution is denoted as $x \sim p_{\text{data}}(x)$ and $y \sim p_{\text{data}}(y)$.

The model consists of two mappings: $G : X \to Y$ and $F : Y \to X$. Additionally, two adversarial discriminators are introduced, $D_X$ and $D_Y$. $D_X$ aimed to distinguish between images $x$ and translated images $F(y)$, while $D_Y$ aimed to discriminate between $y$ and $G(x)$. The objective for $G : X \to Y$ and its discriminator $D_Y$ can be expressed as [19]

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]\end{aligned} \tag{2.12}$$

A similar adversarial loss for $F : Y \to X$ and its discriminator $D_X$ is introduced as well.

The cycle consistency loss in cycleGAN is a crucial component of the objective function that enforces consistency between the original input images and their reconstructed counterparts after undergoing the translation process. It ensures that translating an image from one domain to another ($X$ to $Y$) and then back to the original domain ($X$) results in an image that closely resembles the original input. The cycle consistency loss can be expressed as [19]

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].\end{aligned} \tag{2.13}$$

In some implementations, identity loss [106] is included to preserve the identity of the input images. The identity loss penalizes deviations between the translated image and the original image when both are fed through the generator without any changes. This loss helps maintain the characteristics of the input image during the translation process.

$$\mathcal{L}_{\text{identity}}(G, F) = \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(y) - y\|_1] + \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(x) - x\|_1] \tag{2.14}$$

The overall objective function of cycleGAN combines adversarial loss, cycle consistency loss, and identity loss is given by

$$
\begin{aligned}
G^*, F^* = \arg\min_{G,F} \max_{D_x, D_Y} \mathcal{L}\left(G, F, D_X, D_Y\right) = \mathcal{L}_{\text{GAN}}\left(G, D_Y, X, Y\right) \\
+ \mathcal{L}_{\text{GAN}}\left(F, D_X, Y, X\right) \\
+ \lambda \mathcal{L}_{\text{cyc}}(G, F) \\
+ 0.5\lambda \mathcal{L}_{\text{identity}}(G, F),
\end{aligned}
\tag{2.15}
$$

where the importance of each component in the objective function is determined by weighting them during training using the parameter $\lambda$. Following [19], $\lambda$ is set to 10, while the weight of the identity loss is chosen to be half of the weight assigned to the cycle consistency loss, as recommended in the same reference.

Figure 2.6 illustrates simplified architectures of the pix2pix and cycleGAN models. The distinction in approaches between pix2pix and cycleGAN allows pix2pix to excel in tasks requiring paired training data, where precise mapping is essential. It is well-suited for tasks such as style transfer, semantic segmentation, and image synthesis. In contrast, cycleGAN is more flexible and capable of learning mappings between unpaired domains. This makes it suitable for tasks where paired data are not available.

### 2.1.7 Modified K-means Clustering (K*-means)

In various imaging techniques, time-domain signals are captured at each pixel, offering the possibility of containing significant multimodal information about the observed targets. However, conventional approaches like Principal Components Analysis (PCA) [107, 108, 109] and regular K-means clustering [110, 111, 112] are not well-suited for effectively extracting this information. To address this limitation, Pellegrino et al. [94, 113] introduced a novel clustering method that utilizes an angular distance metric to cluster signals based on their shape rather than amplitude. By calculating centroids as the direction of maximal intra-cluster variance, this method enables the identification of centroids associated with the underlying target characteristics in a scalable and noise-robust manner.

PCA aimed to identify the principal components that explain the maximum variance in the data. However, it does not directly infer meaningful features from the time-domain signals. In the context of imaging modalities like PAM, where the signals can have complex shapes and patterns, PCA may not be well-suited to capture the diverse characteristics and subtle variations present in these signals [113].

On the other hand, regular K-means clustering assumes that each data point belongs to one of a set of prototypes, or clusters. This assumption oversimplifies the inherent

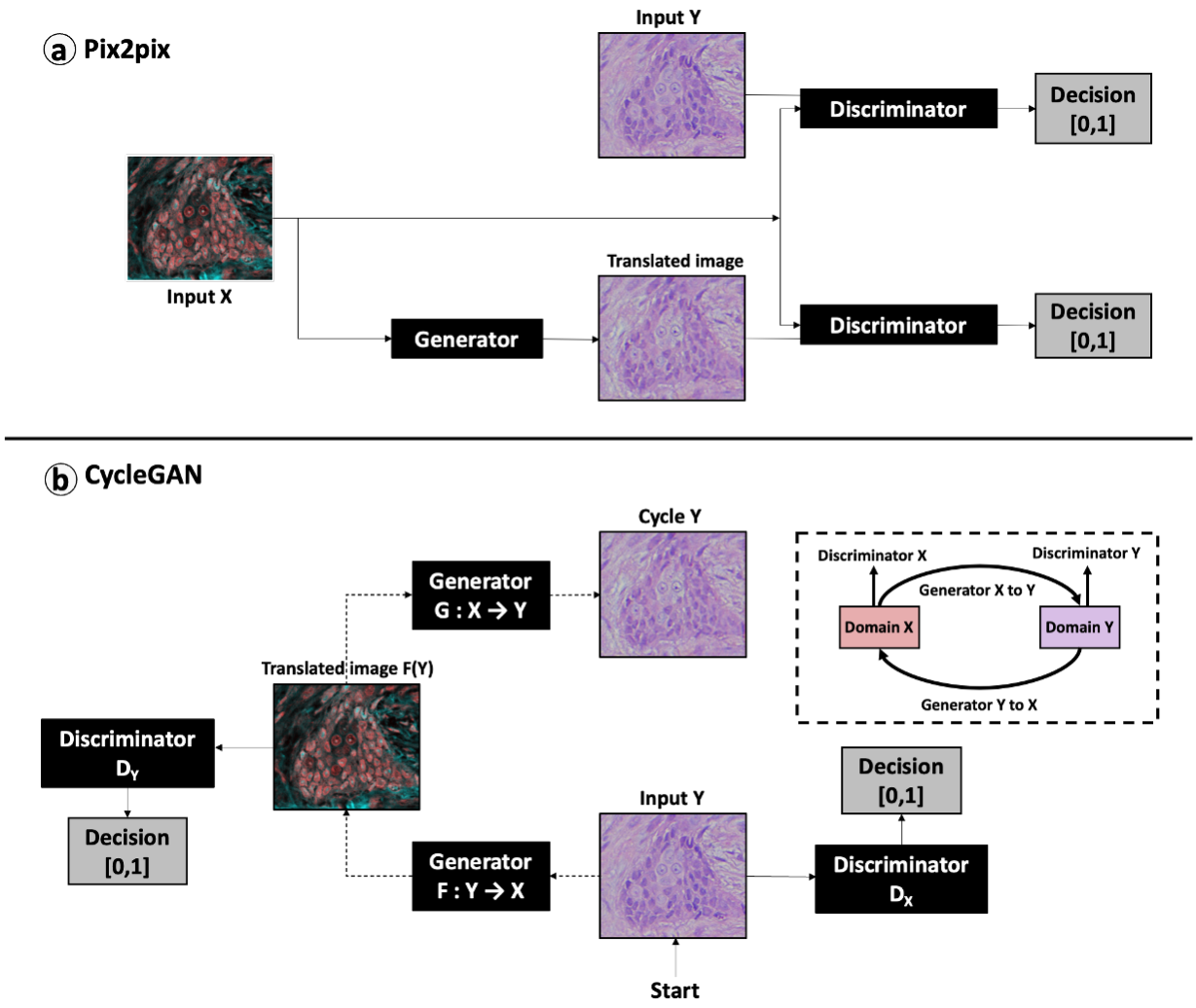



Figure 2.6: The architectures of two image translation models: (a) pix2pix and (b) cycle-GAN.

complexity of time-domain signals, which can exhibit diverse shapes and multimodal patterns [113]. The rigid assignment of data points to a single prototype fails to capture the intricate variations and potential target characteristics embedded within the signals.

To address these limitations, the innovative K*-means clustering method proposed by Pellegrino et al. introduces a different approach that focuses on the shape rather than the amplitude of the signals. By utilizing an angular distance metric, this method captures the underlying structure and shape variations within the time-domain signals. It recognizes that shape information is crucial in identifying and differentiating the various target characteristics that may be present.

In this method, centroids are calculated as the direction of maximal intra-cluster variance. This means that the clustering algorithm not only considers the position of data points in space but also their orientation. By learning centroids based on shape-related characteristics and maximizing the variance within each cluster, this approach provides a more robust and scalable means of clustering time-domain signals.

By leveraging the unique properties of time-domain signals and focusing on their shape information, this novel clustering method has the potential to extract the implicit multimodal information contained within the signals. Therefore, this clustering method offers a promising avenue for effectively uncovering and understanding the complex target characteristics present in imaging modalities like PAM.

Synthetic data experiments were conducted to evaluate the performance of the proposed method in comparison to PCA and Angular-Distance K-means. The results show that the proposed method consistently outperforms the other methods in accurately learning the ground truth prototypes and achieving high centroid similarity [94, 113]. Additionally, the proposed method is applied to real-world data obtained from a PARS microscope. The extracted features demonstrated close correspondence with tissue structures, indicating the potential of the method for capturing biologically meaningful information.

In conclusion, the method proposed by Pellegrino et al. provides a novel approach to clustering time-domain signals by focusing on their shape rather than their amplitude. By leveraging an angular distance metric and computing centroids based on intra-cluster variance, the method effectively captures the underlying signal shapes related to the imaged targets. This approach overcomes the limitations of traditional methods and offers a scalable, noise-robust, and flexible alternative for feature learning and clustering in various imaging modalities. Chapter 5 of the thesis provides a detailed exploration of how this clustering method is applied in the context of virtual staining.

## 2.2 Related Work

In this section, we explore the concept of virtual histology, which involves generating histopathology-like images without traditional staining, using optical microscopes. We delve into a curated collection of deep learning-based virtual staining methods developed for different microscope types.

Several optical microscopes have been developed to provide H&E-like imaging capabilities, such as hyperspectral microscopy [114], Fourier transform infrared (FTIR) spectroscopy [115], bright-field microscopy [116], quantitative phase imaging (QPI) [42], reflectance confocal microscopy (RCM) [9], microscopy with ultraviolet surface excitation (MUSE) [117, 118], optical coherence tomography (OCT) [119, 120, 121], autofluorescence microscopy [13,122], and photoacoustic microscopy (PAM) [58,63,123,124]. These methods are usually coupled with deep learning-based virtual staining models, particularly GANs. Microscopic images are captured and then used to train a network to perform virtual H&E staining. Hence, the quality of the resulting virtual H&E staining will be directly determined by the characteristics of the optical microscope used to capture the images.

In the ideal case, the microscopy technique would (1) operate in reflection mode, enabling imaging of thick tissue specimens, and (2) provide analogous contrast to current histochemical staining, improving artificial colorization quality by avoiding inferences of essential diagnostic features. Other optical microscopes that have exhibited distinguished progress in H&E-like imaging include light-sheet microscopy (LSM) [125] and stimulated Raman scattering microscopy (SRS) [11, 12].

Multiple network architectures have been employed for virtual staining, with GAN emerging as a prevalent and widely embraced framework due to its robust representation capabilities [122]. In contrast to non-GAN inference models, GANs have the capacity to produce images of higher resolution and greater perceptual realism [122]. Given the extensive utilization of GANs in the realm of virtual staining research, our review primarily concentrates on GAN-based frameworks, summarized in Table 2.1.

It is worth noting that cycleGAN and its variants are frequently employed in unsupervised training scenarios [122], particularly in the absence of paired data. However, in situations where precisely registered pairs of input and target images are readily accessible, the adoption of a supervised learning approach is often deemed the most suitable strategy for training a GAN-based virtual staining network [122]. This preference arises from the ability to accurately compute pixel-wise loss functions, thereby facilitating the optimization of the generator.

In instances where paired input and ground truth images are available, conventional

evaluation metrics such as Structural Similarity Index (SSIM) [126], Peak Signal-to-Noise Ratio (PSNR) [127], Multi-Scale Structural Similarity (MS-SSIM) [128], Mean Squared Error (MSE), and Mean Absolute Error (MAE) are commonly employed. Conversely, when paired input and ground truth images are not available, reference-free metrics like the Fréchet Inception Distance (FID) and Inception Score (IS) [129] can serve as viable alternatives for the purposes of evaluation. Below, we investigate the realm of virtual histology across a range of optical microscope technologies.

### Hyperspectral imaging

Hyperspectral imaging [114] is an advanced imaging method that captures a wide range of spectral information from a scene or object, going beyond traditional color imaging. It records numerous narrow spectral bands, allowing for detailed material analysis and finding applications in various fields like remote sensing, agriculture, and medical imaging. In the context of virtual staining, a pioneering research paper [130] introduced a novel approach for virtually staining unstained tissue specimens using dimension reduction and cGANs. The study focuses on capturing unstained tissue samples through a hyperspectral camera and leveraging dimensionality reduction techniques and cGANs to transform these images into virtually stained versions. The authors aim to overcome the challenges associated with traditional chemical staining, such as variability in appearance and adverse effects on tissue structures.

Table 2.1: Summary of GAN-based virtual staining frameworks using various imaging modalities.

| **Imaging Modality** | **Deep learning Model** | **Registration Requirement (Y/N)** | **Evaluation Method** | **Modality Limitation** |
|---|---|---|---|---|
| Hyperspectral Imaging | cGAN (2017) [130] | Y | SSIM, MSE | Cost, complexity |
| | Pix2pix, CycleGAN, CUT (2023) [131] | Y, N, N | SSIM, Chromatic discrepancy + Blind id of stain type | |
| Bright-Field Microscope | Cond. cycleGAN (2019) [132] | N | Expert/Non-expert eval. | Low contrast/specificity |
| | cGAN (2020) [116] | Y | Relative error + Blind eval. | |
| | Pix2pix (2021) [133] | Y | SSIM, PSNR, LPIPS + Blind eval. | |
| | Self-attention pix2pix (2021) [134] | Y | SSIM, PSNR, LPIPS | |
| | StainGAN (2022) [134] | Y | SSIM, UQI + Expert eval. | |
| Fourier Transform Infrared | CNN (2019) [115] | Y | SSIM | Unsuitability for thick tissue |
| Quantitative Phase Imaging | PhaseStain (2019) [42] | Y | SSIM | Unsuitability for thick tissue |
| Autofluorescence Microscopy | GAN (2019) [13] | Y | Blind eval. | Low chromophore specificity |
| | GAN (2022) [135] | N | FID | |
| Microscopy with Ultraviolet Surface Excitation | Pix2pix, cycleGAN (2021) [117] | Y, N | PCC, SSIM, MMD, PSNR | Reliability on exogenous dyes |
| Reflectance Confocal Microscopy | Pix2pix (2021) [9] | Y | PCC, SSIM, Morphological metrics | Unsuitability for thick tissue |
| Optical Coherence Tomography | Pix2pix (2021) [5] | Y | Double-blind eval. | Low chromophore specificity |
| Photoacoustic Microscopy | CycleGAN (2022) [8] | N | Morphological metrics | Reliance on contact-based detection |

The virtual staining approach proposed by Bayramoglu et al. has the potential to automate tasks in diagnostic pathology, improve processing time, reduce costs, and increase objectivity in disease diagnosis. The authors emphasize the need for validation by pathologists and plan to evaluate the results through pathologists' feedback in the future. If clinically accepted, virtual histology staining could provide an objective diagnosis, eliminate visual variabilities, and streamline the diagnostic pathology workflow.

Another very recent work by Salido et al. [131] focused on the quantitative comparison of three generative models of H&E virtual staining applied to hyperspectral images of breast tissue samples. The models, based on cGAN, cycleGAN, and contrastive learning, were evaluated based on structural similarity and chromatic discrepancy compared to chemically stained samples. The results (illustrated in Figure 2.7) showed that cycleGAN demonstrated exceptional performance in relation to structural similarity and lower chromatic discrepancy. The authors also conducted a qualitative evaluation of the results using subjective psychophysical tests with experts, further validating the quality of the best model. The study highlights the importance of generative staining models that guarantee cyclic consistency in producing results closest to chemical H&E staining.

### Bright-Field Microscope

The primary drawback of hyperspectral imaging is its bulkiness, cost, and complexity. In contrast, bright-field microscopy stands out as one of the simplest microscope techniques. In an effort to address the challenges associated with variable and time-consuming tissue staining procedures in histological analysis, a recent study [116] introduced a deep learning approach based on cGAN. The researchers trained a convolutional neural network to transfer unlabeled bright-field microscopic images of tissue sections into equivalent images that resemble histologically stained versions of the same samples. The results of blind evaluations by pathologists demonstrated that there were no significant differences between the virtual staining and standard histological staining images, even when different types of stains were used. Quantitative analysis of carotid artery tissue sections revealed minimal relative errors in the measured parameters.

Zhang et al. [133] introduced a deep learning-based method for virtually staining human carotid atherosclerotic plaques, addressing the limitations of traditional histological staining techniques. By utilizing a pix2pix model, the authors achieve simultaneous image-to-image translation of bright-field images of unlabeled tissue sections into multiple virtually stained equivalents. The proposed approach demonstrated superior performance compared to existing methods, as validated by evaluation metrics and blind evaluations by pathologists. It accurately captures key histopathological features of atherosclerotic

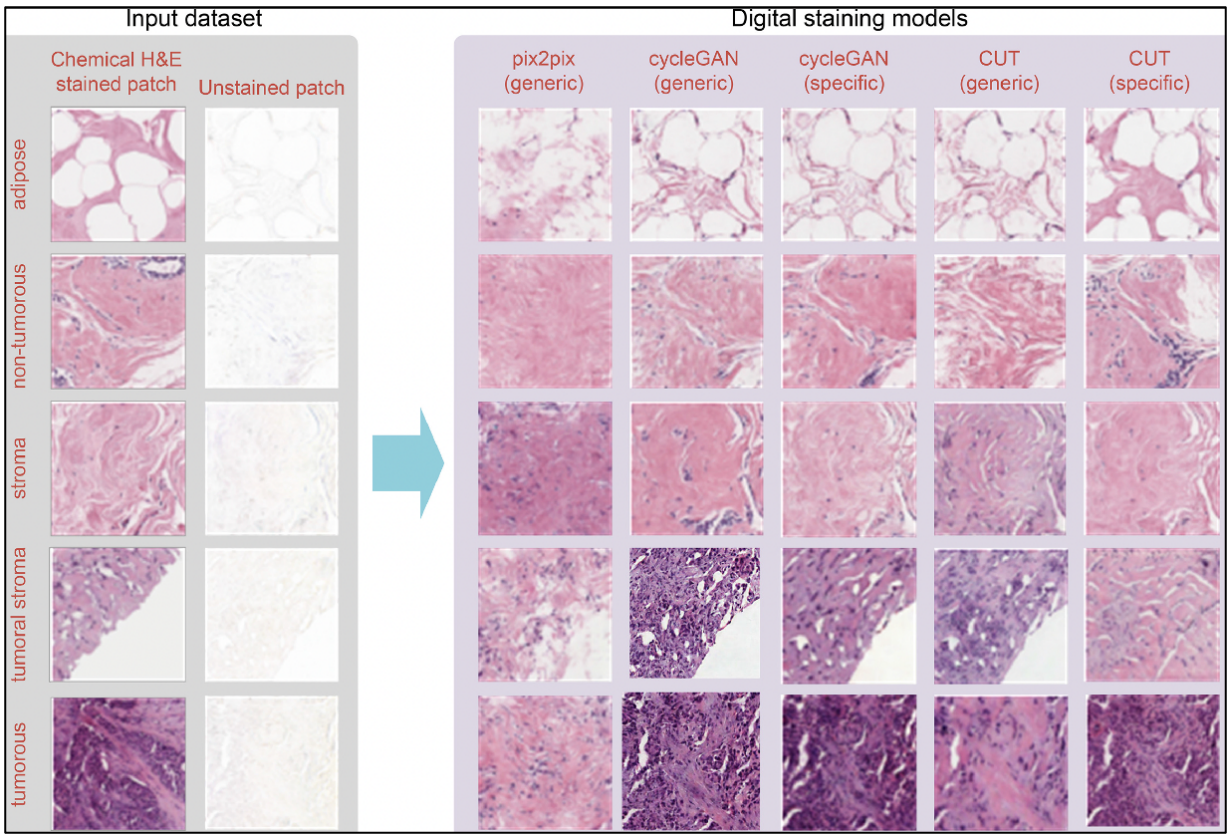



Figure 2.7: The results of digital staining using both generic and tissue-specific models were evaluated based on input images, including target chemical staining images and unstained images. Source: Reprinted from [131].

plaques and enables virtual staining of unlabeled tissue samples, offering new opportunities for personalized prevention and interventional treatments in carotid atherosclerosis. This advancement not only improves staining efficiency and consistency but also enhances our understanding of atherosclerosis biology, leading to more effective strategies for plaque prevention and personalized treatment of ischemic stroke.

Another work by Zhang et al. [136] proposed a novel framework called Multiple Virtual Functional Stain (MVFStain), which aimed to address the need for artificial intelligence stain generation in histopathology images. While traditional pathological examinations rely on H&E staining for histomorphological characteristics, additional functional staining is often required for an accurate disease diagnosis. However, performing multiple stains on the same tissue section is time-consuming and leads to morphological misalignment.

The MVFStain framework addresses these challenges by producing virtual functional staining images from H&E-stained images. It utilizes domain mapping and incorporates multiple sources of staining information to achieve the desired outcomes. The framework comprises content and style encoders, a decoder, and two discriminators. By separating content and style features, MVFStain facilitates the conversion from H&E staining to various functional stains.

The proposed method is evaluated on four datasets, involving the translation of H&E staining to nine other functional stains. The results demonstrated that MVFStain performs closely to, or even surpasses, one-to-one image translation in terms of image quality metrics. Positive signal prediction and lesion detectability were also evaluated, and the virtual staining shows promising results in terms of accuracy and detectability. The virtual staining results are shown in Figure 2.8. The study highlights the potential of developing MVFStain as a clinical alternative to artificial functional stains, providing robust support for histological analysis and computer-aided diagnosis research. This framework holds promise for expanding the modality of H&E staining and facilitating multimodal pathological diagnosis systems. Further research aimed to improve the accuracy of positive signal prediction and enhance the generality of stain mapping for uncommon functional stains.

In a related investigation conducted by researchers, a notable alternative approach is presented by Zhang et al. [134]. A self-attention GAN model is introduced to virtually stain human carotid atherosclerotic plaque tissue sections, enabling the identification of histopathological features without the need for traditional staining. The proposed approach achieved superior performance compared to other methods, as evaluated by quantitative metrics. The authors believe that this virtual staining method has important clinical relevance as it allows for the analysis of unlabeled tissue samples, facilitating the development of personalized prevention and treatment strategies for carotid atherosclerosis. The study

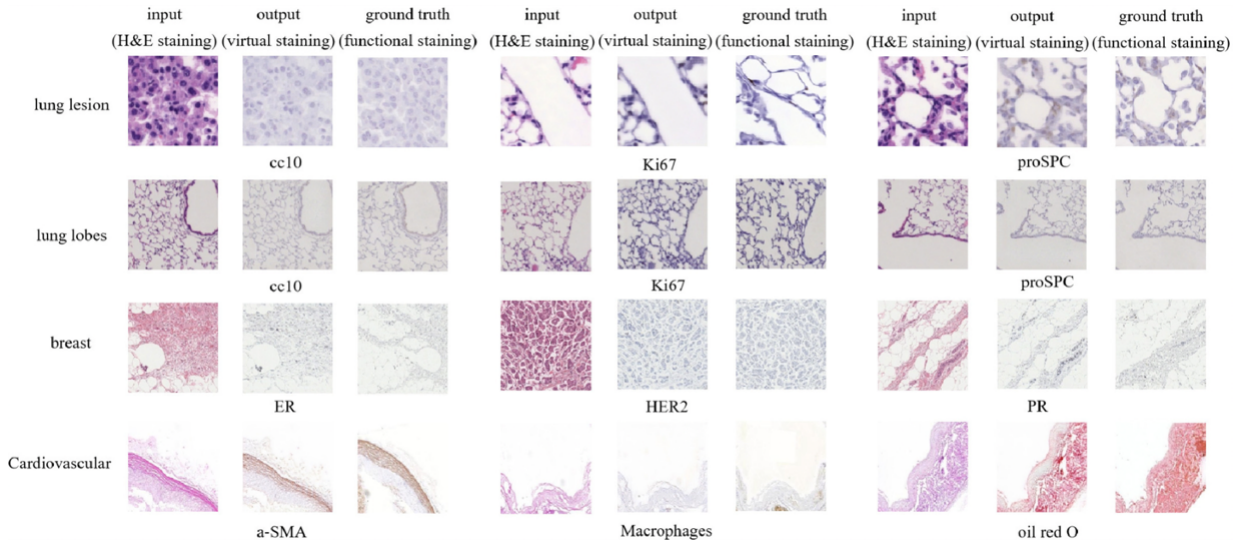



Figure 2.8: Experimental verification of the proposed network using four datasets: lung lesion, lung lobes, breast, and atherosclerotic lesion (from H&E stain to other stains). Source: Reprinted from [136].

acknowledges some limitations, including the small training dataset and the need for further verification and improvement of the method's efficiency. Nevertheless, the adoption of deep learning-based virtual staining offers the potential to overcome the laborious and variable nature of traditional staining processes, enhancing the standardization and efficiency of histological image analysis systems for atherosclerotic plaque evaluation.

Traditional staining methods like H&E provide morphological information, but additional staining is often required for specific protein visualization. However, restaining tissue samples is time-consuming and costly. To address this limitation, a deep learning-based method called StainGAN was introduced [137]. StainGAN utilizes self-attention GAN with contrastive learning and a color-based loss function to transform stained images virtually. The performance of StainGAN is compared to other methods including pix2pix and cycleGAN, and it demonstrated superior qualitative and quantitative results, particularly in producing virtual CD8-stained images, as illustrated in Figure 2.9. StainGAN offers a time-efficient and accurate stain transformation solution for histopathology images, leveraging self-attention and contrastive learning techniques.

While StainGAN shows promising results, challenges remain in accurately estimating T cell density in synthetic CD8-stained whole slide images (WSIs). The low color intensity of immune cells and residual artifacts pose difficulties for robust segmentation

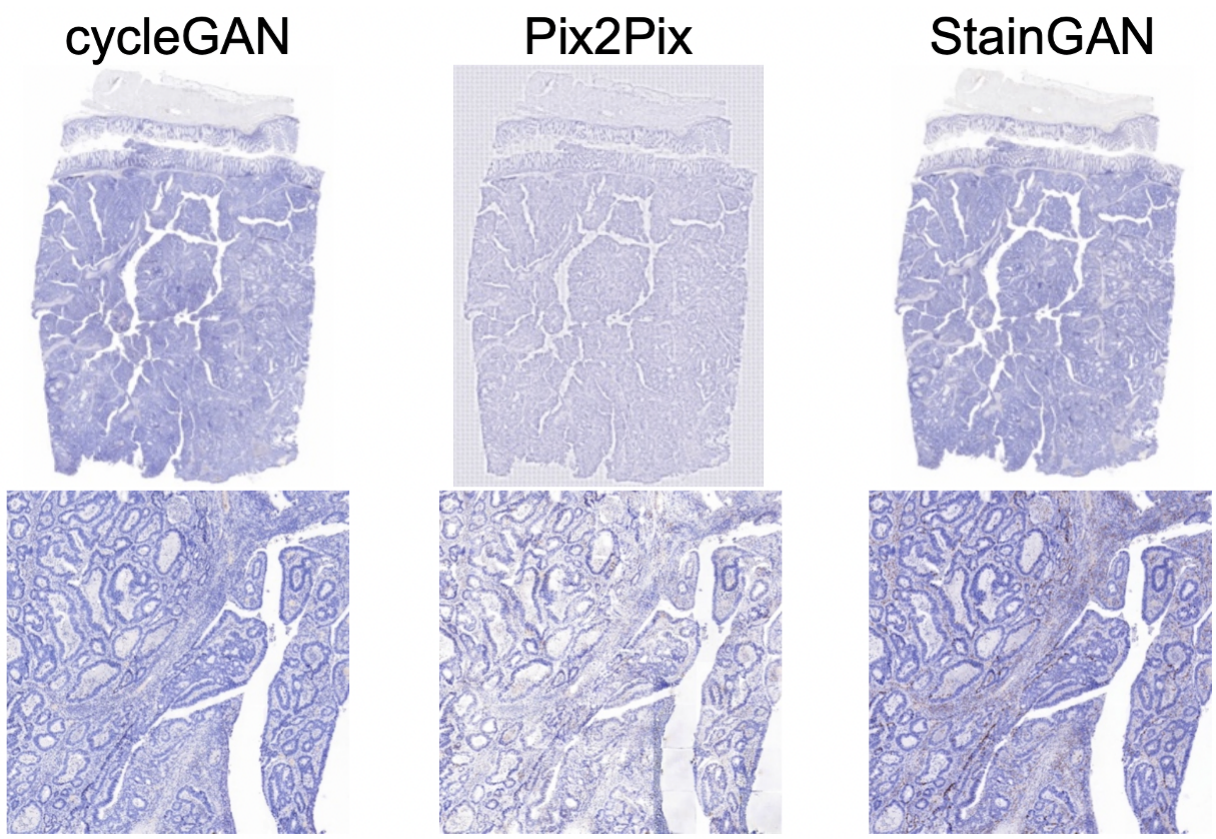



Figure 2.9: The StainGAN results were compared with the cycleGAN and pix2pix models. The first row of the comparison represents whole-slide images (WSIs), while the second row displays patches from these WSIs. Source: Reprinted from [137].

methods. Expanding the dataset used in the study could further enhance the results. In summary, StainGAN combines the strengths of self-attention blocks and contrastive learning to achieve fast and accurate stain transformation, making it a valuable tool in the computer-based analysis of histopathological images.

Another technique was proposed as a solution to tackle the drawbacks of traditional complex and costly immunohistochemistry (IHC) staining methods [132]. A conditional cycleGAN (cCGAN) network is proposed to transform H&E-stained images into IHC-stained images, allowing virtual IHC staining on the same slide. The cCGAN model improves translation accuracy using category conditions and structural loss functions. This approach offers an alternative to traditional staining by using deep learning to generate realistic stained-like images, providing advantages in time, cost, and staining variability. However, the widespread adoption of bright-field microscopy was hindered by the microscope's low contrast, posing limitations. To compensate for this, the colorization network needs to estimate features that cannot be directly measured.

### Microscopy with Ultraviolet Surface Excitation (MUSE)

Chen et al. [117] have proposed a new workflow for histopathology using the MUSE imaging technique combined with deep learning algorithms. This workflow aimed to provide rapid and accurate histological images without the need for lengthy sample preparation. By translating colorful MUSE images into Deep-MUSE images that resemble traditional H&E staining, the authors simplify the histopathology process.

The MUSE imaging system allows for the imaging of fresh and thick tissues with specific molecular contrast, eliminating the need for thin tissue slices. The proposed workflow uses an unsupervised GAN to convert MUSE images into Deep-MUSE images that closely resemble H&E-stained images. The results are depicted in Figure 2.10. This approach significantly reduces the time required to obtain histological images, with a typical brain biopsy image acquired in 5 minutes and translated into a Deep-MUSE image in just 40 seconds.

The advantages of this workflow lie in its speed and consistency. Deep-MUSE images can be acquired rapidly, and the deep learning model generates virtual H&E-stained images with a consistent color tone, reducing variations caused by different staining conditions. The proposed workflow holds great promise for intraoperative histopathology, providing a fast and accurate assessment tool for surgeons and pathologists. Further research is needed to test the accuracy of the workflow on human specimens and explore its potential applications to different organs and staining methods.

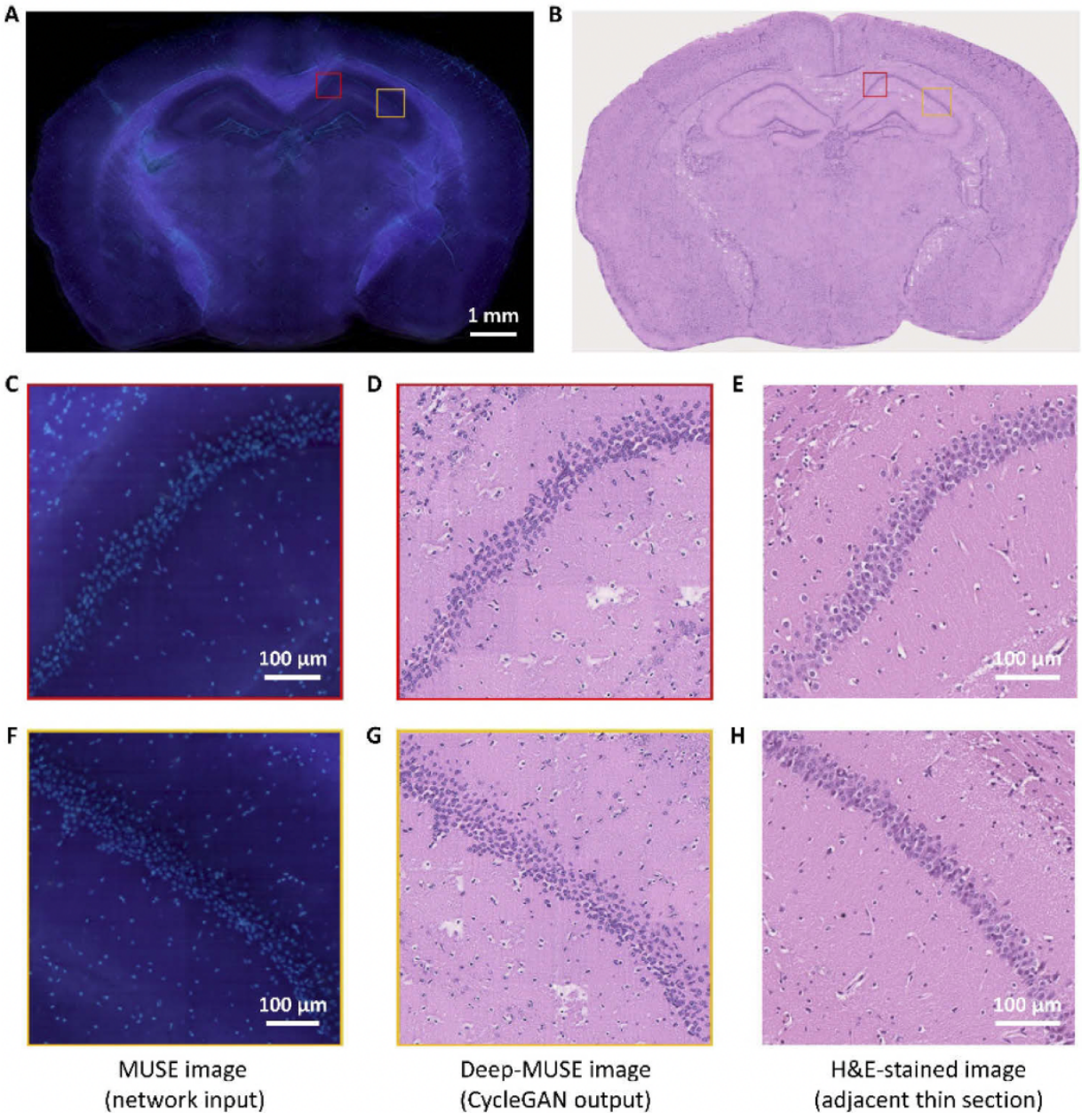



Figure 2.10: A comparison between MUSE and Deep-MUSE images of a fresh mouse brain, alongside H&E-stained images of the adjacent thin section. Source: Reprinted from [117].

Although MUSE and another technique, namely LSM, have demonstrated acceptable agreement with H&E-stained histology [117, 125], mimicking the standard H&E staining specificity, both techniques rely on fluorescence stains to provide H&E contrast. These dyes can potentially be mutagenic, carcinogenic, or toxic, hindering the in-situ applicability of these fluorescence-based techniques [118]. Moreover, LSM requires relatively clear samples for 3D volumetric imaging of fresh tissue, which necessitates additional equipment and long preprocessing times making the technique impractical in an intraoperative setting [125]. Accordingly, label-free techniques are expedient.

### Reflectance Confocal Microscopy (RCM)

Other imaging techniques including RCM, FTIR, and QPI have proven successful in generating virtual stains, yet less suitable for thick tissue imaging due to their mode of operation. Li et al. [9] presented a novel deep learning-based framework for virtual histology that enables the rapid transformation of label-free in vivo RCM images of unstained skin into virtually stained images resembling H&E stained histology. The framework utilizes a convolutional neural network trained under an adversarial learning scheme to generate microscopic-resolution images of normal skin, basal cell carcinoma (BCC), and melanocytic nevi with pigmented melanocytes.

In this research paper, the label-free, unstained image stacks were accurately registered to the corresponding acetic acid-stained image stacks. This registration process provided a high degree of spatial supervision for the neural network to map label-free RCM images to their histological counterparts. By aligning the unstained and stained image stacks, the network was able to learn the image-to-image transformation for virtual staining.

The virtual staining results (illustrated in Figure 2.11) demonstrated similar histological features to traditional H&E histology from the same excised tissue. The study shows that deep learning-based virtual staining can be applied to non-invasive imaging technologies like RCM, potentially enabling more rapid diagnoses of malignant skin neoplasms and reducing unnecessary invasive skin biopsies.

The authors highlight the limitations of existing non-invasive optical imaging technologies for diagnosing skin tumors, such as RCM, due to their grayscale nature and lack of nuclear features. The proposed framework addresses these limitations by virtually staining the RCM images, allowing visualization of the epidermis, dermal-epidermal junction, and superficial dermis layers. The results demonstrated good concordance between the virtually stained images and conventional histology. The virtual staining approach has the potential to transform non-invasive imaging technologies by providing a virtual "optical

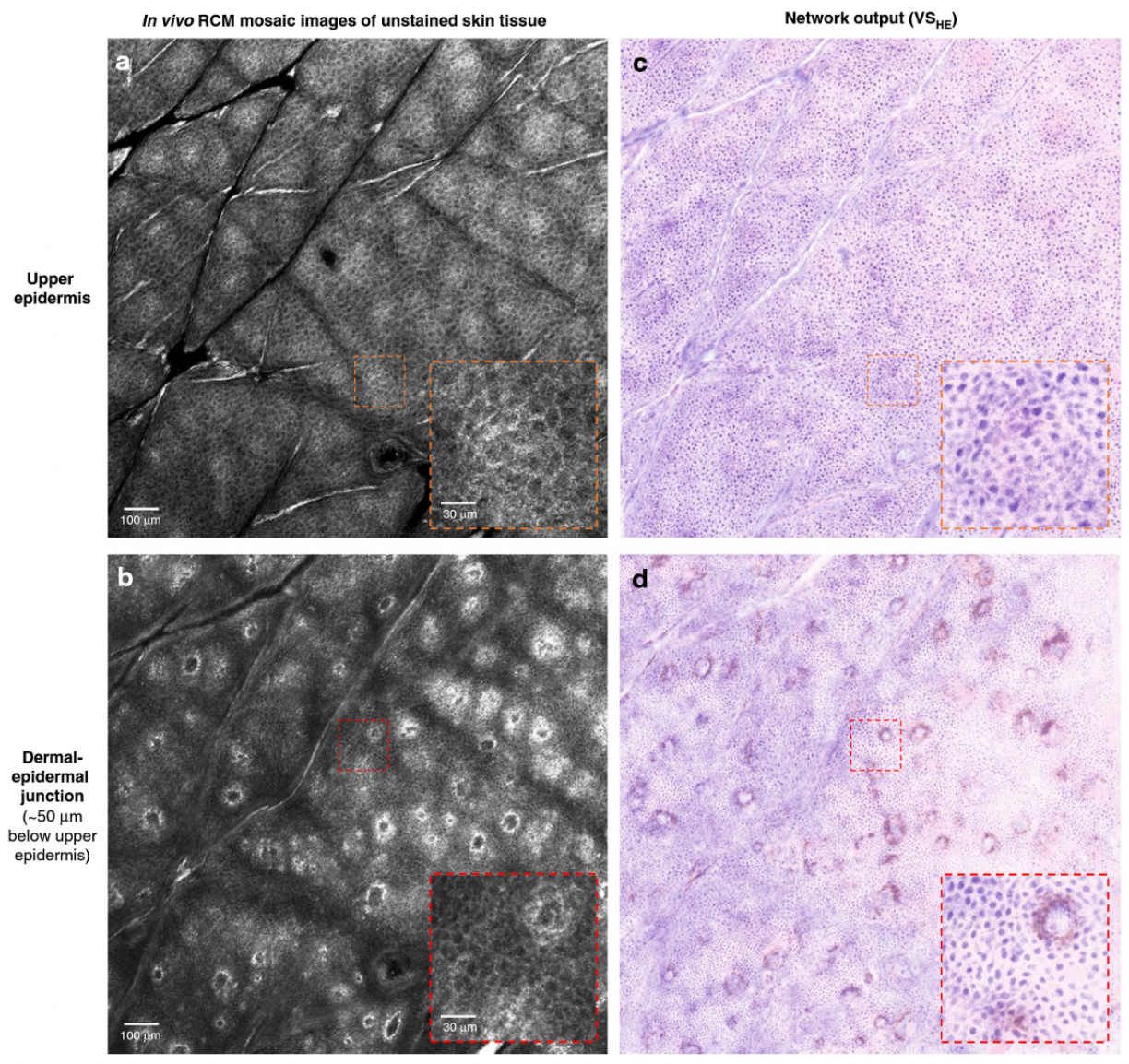



Figure 2.11: The results of pseudo-H&E virtual staining are presented for large field-of-view mosaic images of in vivo skin tissue at two different depths. Source: Reprinted from [9].

biopsy" with cellular resolution and an easy-to-interpret visualization format. However, the authors acknowledge the need for further research and larger datasets to evaluate the utility and diagnostic accuracy of the virtual histology algorithm.

### Fourier Transform Infrared (FTIR)

FTIR spectroscopic imaging [115] has demonstrated potential for characterizing significant tissue phenotypes without the need for chemical labels. However, current FTIR classification methods often rely on laborious and error-prone human annotation in supervised learning approaches. To address this, digital staining using machine learning has emerged as an alternative, enabling the mapping of infrared spectra to corresponding chemical stains, and offering a more efficient and accurate labeling process.

Lotfollahi et al. (2018) [115] introduced an early instance of virtual staining that relied on deep learning. Their approach involved the utilization of FTIR spectroscopy in conjunction with deep learning techniques to digitally stain histological images. The proposed approach aimed to overcome the limitations of traditional histological stains by mapping FTIR images to images that mimic the appearance of chemically stained H&E tissue. The authors demonstrated that high-resolution mappings can be achieved using FTIR imaging and histological staining of the same sample, as depicted in Figure 2.12. They employed convolutional neural networks (CNNs) that leverage both spatial and spectral features to improve the quality of the virtual staining. The method shows promising results, offering cellular-level resolution and providing reliable representations of tissue for pathological inspection. Further validation and exploration of alternative contrast mechanisms are suggested by Lotfollahi et al. for future research.

### Quantitative Phase Imaging (QPI)

Using a deep neural network called PhaseStain, Rivenson et al. [42] have developed a digital staining technique that can transform QPI images of label-free tissue sections imaged using a holographic microscope into images that resemble bright-field microscopy images of the same samples after histological staining. The PhaseStain network is trained using paired image data, and it effectively performs virtual staining on human skin, kidney, and liver tissue samples, matching the staining styles of H&E, Jones' stain, and Masson's trichrome stain, respectively. This digital staining approach eliminates the need for traditional histological staining, reducing costs and saving time. The results demonstrated the potential of data-driven image transformations facilitated by deep learning in pathology applications and biomedical research.

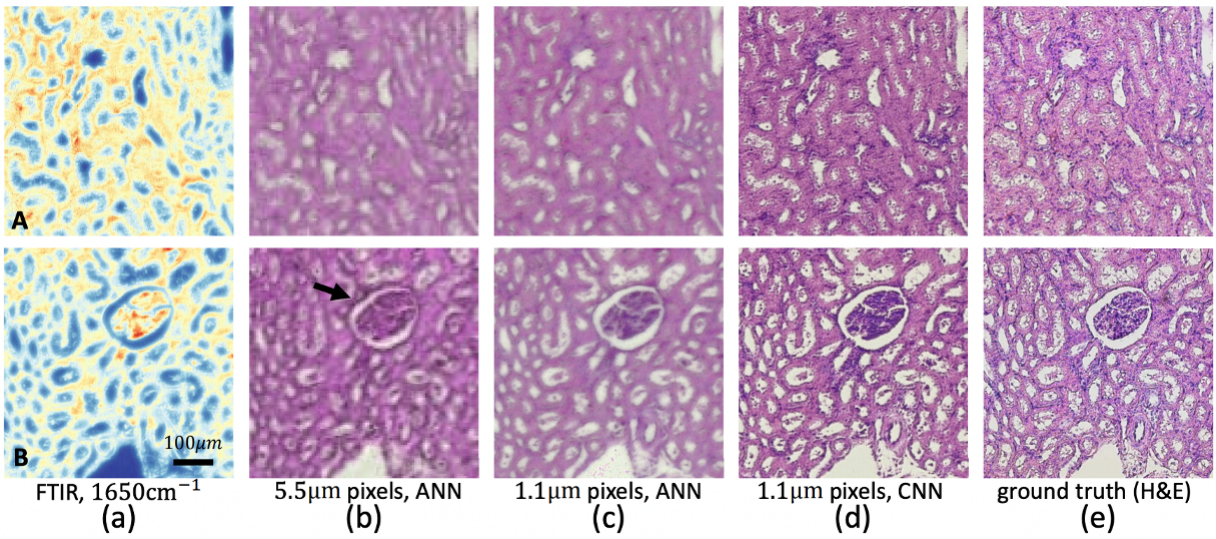



Figure 2.12: Molecular imaging achieved through chemical imaging using various digital staining methods and different spatial resolutions. The figure showcases different image representations: (a) FTIR images, depicted as a single band, (b) Computed H&E images acquired via an Artificial Neural Network (ANN) using low magnification data, (c) Computed H&E images obtained using the ANN framework with high magnification data, (d) Computed H&E images obtained by a Convolutional Neural Network (CNN) with high magnification data, (e) physically-stained H&E images serving as the ground truth. Source: Reprinted from [115].

The PhaseStain network operates at a fast computation speed and can be integrated with a QPI-based whole slide scanner, enabling the generation of virtually stained images while the tissue is being scanned. By circumventing the staining process, this technology not only saves time and costs but also preserves the tissue for other analyses such as molecular and genetic studies. The study demonstrated the feasibility of training the network for specific stain types and tissue combinations, and the PhaseStain framework can be applied to various QPI techniques used in coherent imaging systems. Coherence-related image artifacts can be addressed through additional deep learning-based methods to enhance the quality of digitally stained images.

### Optical Coherence Tomography (OCT)

Another technique, namely OCT, usually operates in reflection mode and is well suited to imaging thick tissue specimens. OCT has shown potential in label-free tissue block imaging [120], surgical margin assessment [138, 139], and biopsy examination [140, 141].

A technique called "optical barcoding" was developed by Winetraub et al. [5] that allowed them to synthesize virtual H&E tissue sections from OCT images. By training a cGAN on a dataset of aligned OCT and H&E images, they were able to generate high-fidelity virtual H&E sections and 3D volumes, as presented in Figure 2.13. This approach has the potential to enable physicians to incorporate OCT imaging into clinical practice, reducing the need for invasive biopsy procedures. The researchers also discussed future directions for their work, including using optical barcoding to obtain precise ground truth for OCT images and applying the OCT-to-Histology process for various research and clinical applications, such as diagnosing skin cancers and confirming tumor margins non-invasively.

The authors acknowledged some limitations of their machine-learning algorithm. They trained their model on a relatively small dataset, but they expect the model to improve as more data becomes available. OCT images have limitations in resolution, signal quality, and contrast, which affect the level of detail that can be reliably generated by the neural network. The authors focused on generating larger features in the OCT images and regarded fine histology features as general texture features rather than precise cell positions. They believe that higher-resolution OCT systems in the future will produce virtual H&E images of higher quality. Overall, the researchers believe that their OCT-to-Histology model has the potential to support clinical decision-making in dermatology and provide valuable insights into the relationship between different optical imaging methods and histopathological sections.

Despite the efforts, the optical scattering contrast used in OCT primarily emphasizes

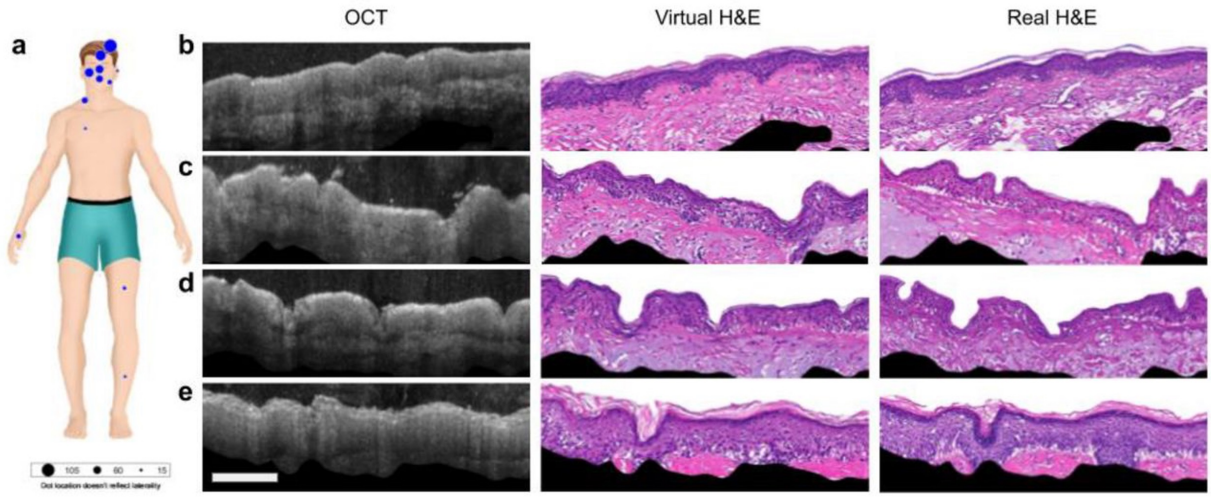



Figure 2.13: Virtual H&E histology generation from OCT images. Source: Reprinted from [5].

morphological details and does not provide enough specificity to differentiate biomolecules such as cell nuclei and cytoplasm, which are critical for histological analysis. Subsequently, artificial colorization methods must infer the structures of these essential diagnostic features, which greatly reduces diagnostic confidence. This motivates the adoption of a label-free microscopy modality with higher chromophore specificity, which may improve colorization confidence.

### Autofluorescence Microscopy

Recently, autofluorescence microscopy [13] has emerged as a popular contender for virtual histology. Autofluorescence provides label-free visualizations of biomolecules like elastin, collagen, amino acids, NADPH, and other cellular organelles, at excitation wavelengths ranging from UV to ∼500 nm [142]. In a groundbreaking contribution to digital pathology, a recent study (2019) [13] presented a novel technique for virtually staining tissue samples using autofluorescence images, employing a convolutional neural network (CNN) approach. The authors validated the efficacy of their method by conducting a comparative analysis between virtual staining and conventional histological staining across different tissues and stains, revealing a lack of significant discrepancies.

By training a CNN using GAN models, the authors achieved a transformation from label-free autofluorescence images to bright-field images that were equivalent to histolog-

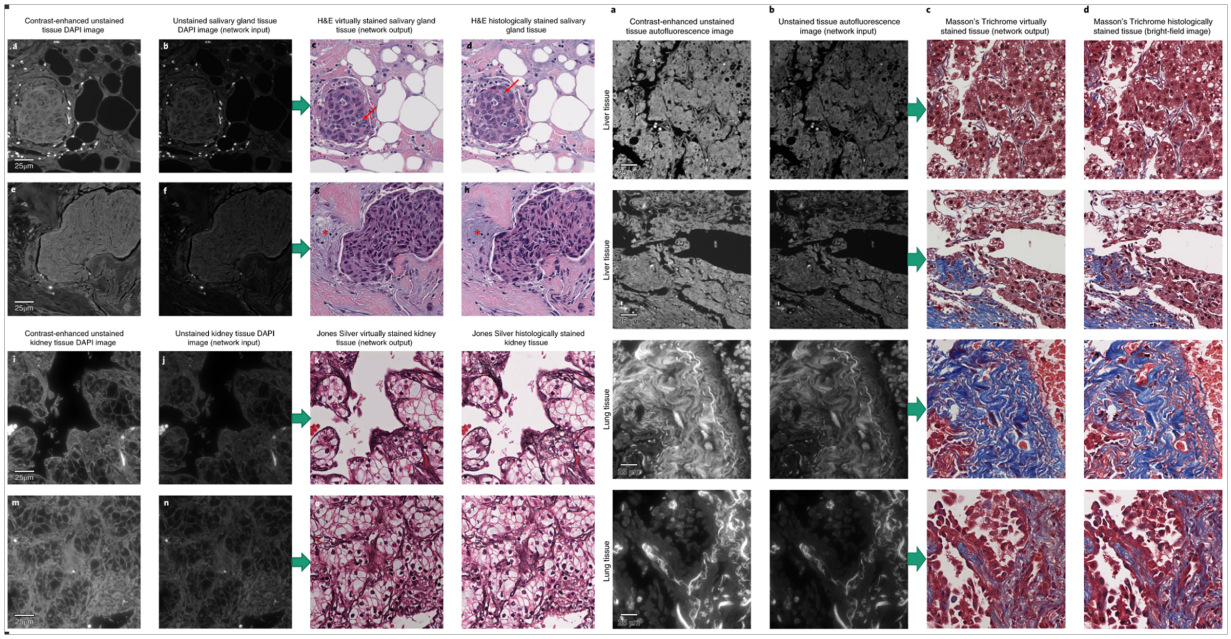



Figure 2.14: Virtual staining of unlabeled autofluorescence images of salivary gland tissue samples (left) and liver and lung tissue samples (right). Source: Reprinted from [13].

ically stained samples, as illustrated in Figure 2.14. This approach has the potential to streamline the clinical workflow in histopathology and can be extended to other label-free imaging modalities beyond autofluorescence.

One important challenge the authors faced was the training process, which involved matching autofluorescence images of unlabeled tissue samples with corresponding stained images. The authors address this by implementing a multi-stage registration process and setting a threshold for acceptable correlation values to ensure high-accuracy local registration. They also discuss the potential for combining multiple imaging modalities and excitation wavelengths to enhance the inference of different tissue constituents. Additionally, the paper highlights the need for further validation through clinical studies to assess the diagnostic accuracy of the virtual staining method and explores potential applications, such as guiding molecular analysis and personalized therapy development.

Furthermore, De Haan et al. [143] demonstrated the effectiveness of a supervised learning-based computational stain transformation technique from H&E stain to different special stains using kidney tissue sections. The authors show that virtual special stains generated from existing H&E images improve the diagnosis of non-neoplastic kidney

diseases. The quality of the transformed special stains is comparable to those produced through standard histochemical staining. The approach offers advantages such as reduced slide preparation time, decreased laboratory costs, and tissue preservation for subsequent analysis. The stain transformation network is trained using pairs of perfectly registered training images created by label-free virtual staining of autofluorescence images, ensuring reliable and accurate transformation. The method also generalizes well across stain variations and can be applied to other types of stains. However, further validation with larger datasets and different imaging setups is necessary for diagnostic use. The technique complements standard light microscopy evaluation and may be extended to other biomarkers or imaging modalities, enhancing diagnostic capabilities and saving resources. The stain transformation process is rapid and computationally efficient, benefiting both patients and the healthcare system.

Another paper published by the same research group, entitled "Virtual Stain Transfer in Histology via Cascaded Deep Neural Networks," [144] introduced a novel framework for digitally transforming H&E-stained tissue images into other histological stains using deep neural networks. The authors address the challenge of obtaining multiple stains on the same tissue section by proposing a cascaded deep neural network (C-DNN) approach. This method first virtually stains autofluorescence microscopy images into H&E and then performs stain transfer from H&E to the target stain.

To validate their approach, the authors, Yang et al., used kidney needle core biopsy tissue sections and successfully transferred H&E-stained images into virtual PAS stain. The C-DNN approach outperforms a standard stain transfer network in terms of image quality, color distribution, and contrast. By leveraging a cascaded structure and training the network with two groups of data containing different stain types, the authors achieve improved image quality and color accuracy in the virtual stain transfer process.

To train the C-DNN framework, two groups of data were utilized. Group A consists of histochemically stained H&E images and virtually stained PAS images, while Group B consists of histochemically stained PAS images and virtually stained H&E images. By using these two groups of data, the C-DNN framework can directly minimize the loss between the output images and the histochemically stained target images in both the H&E and target stain domains during training.

In another research endeavor [135] focused on virtual histological staining, the aim was to develop a technique for visualizing complex biological structures using label-free autofluorescence images. Previous methods have relied on two-side image translation architectures, which require paired data and suffer from calculation inefficiencies and incorrect staining. In this study, a novel one-side image translation framework is proposed to trans-

form unstained auto-fluorescence images into virtual H&E-stained counterparts for both thin and thick human samples. To address the issue of incorrect nuclear staining, a region classification loss is designed to incorporate partial supervision information. Experimental results using thin and thick human lung samples demonstrated that the proposed method is computationally efficient and achieves comparable transformation performance to the two-side framework. This approach offers a potential alternative for rapid pathology, as stained images can be obtained within minutes.

In spite of these endeavors and the numerous positive results achieved, the broad realization of autofluorescence microscopy in bulk tissues was limited by the autofluorescence contrast. Though autofluorescence highlights a number of biomolecules resembling eosin staining, the DNA and nuclei [145], do not exhibit any measurable autofluorescence [43]. Hence, while many diagnostic elements are present, the colorization network must estimate the nuclear features since they cannot be measured. Inferring nuclear structures (a key determinant of cancer diagnosis) presents challenges to clinical adoption and regulatory approval.

### Photoacoustic Microscopy (PAM)

One modality, PAM, presented in Section 2.1.3, has gained attention for providing selective biomolecule contrast without the need for staining. PAM uses specific excitation wavelengths to visualize cellular structures analogous to traditional hematoxylin and eosin staining. Previously, PAM systems have been shown to recover both nuclear visualizations analogous to hematoxylin staining [146, 147] and connective tissue imaging analogous to eosin staining [123].

Several PAM techniques incorporate virtual histology to enhance the interpretation of the acquired images by pathologists and researchers [8, 148]. One notable milestone in this field is the integration of ultraviolet photoacoustic microscopy (UV-PAM) with deep learning, referred to as Deep-PAM, which enables the generation of virtually stained histological images (depicted in Figure 2.15 [8]. This innovative approach, termed Deep-PAM, offers a rapid and non-destructive method for imaging thin sections, thick fresh tissue specimens, and even intact specimens. The non-destructive nature of UV-PAM allows for the reuse of imaged specimens for further testing, saving valuable resources and reducing costs associated with tissue processing and staining.

Compared to fluorescence-based histological imaging methods, Deep-PAM relies on intrinsic biomolecular contrast, resulting in more consistent images. It eliminates the challenges associated with dye concentration, labeling time, tissue microenvironments,

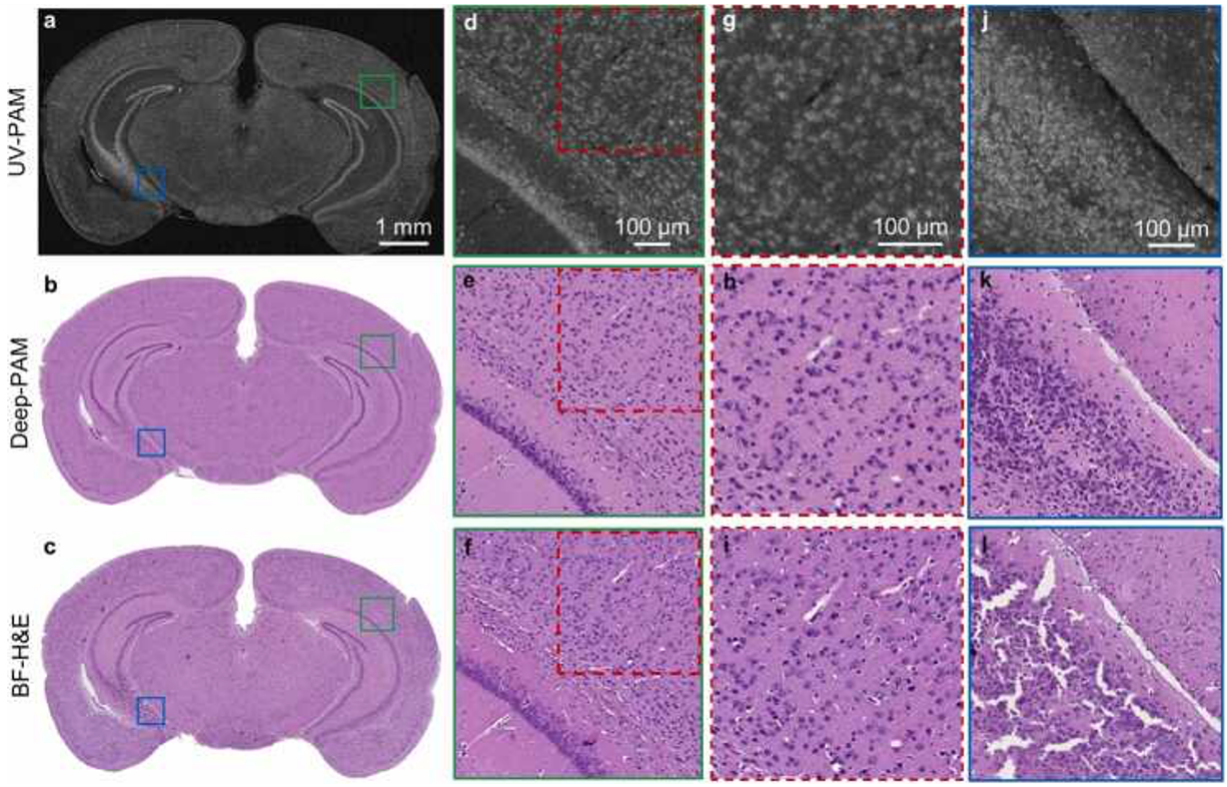



Figure 2.15: The validation results of UV-PAM and Deep-PAM using 7 $\mu m$ deparaffinized FFPE mouse brain section. Source: Reprinted from [8].

and various tissue types. Furthermore, the authors believe that Deep-PAM offers advantages over other label-free imaging modalities such as SRS and Simultaneous label-free autofluorescence-multi-harmonic (SLAM), as it provides higher imaging speed and is more cost-effective and compact.

The generation of virtually stained images in Deep-PAM is achieved through deep learning techniques, particularly the use of unsupervised learning methods like cycleGAN. Unlike previous model-based mapping algorithms, cycleGAN allows for high-quality style transformation between UV-PAM images and bright-field H&E-stained images without the need for paired training data. This eliminates the complexity of image registration procedures and enables the virtual staining of thick samples. The versatility of Deep-PAM extends beyond H&E staining, as it can be applied to other special stains with sufficient training data.

In summary, the integration of UV-PAM with deep learning techniques in Deep-PAM offers a comprehensive and efficient approach to histological imaging. It provides rapid and label-free virtual staining of tissue specimens, enabling preoperative, intraoperative, and postoperative disease diagnosis. The advantages of Deep-PAM over fluorescence-based imaging methods, such as consistent image quality and cost-effectiveness, make it a promising tool in the field of histopathology.

However, PAM's reliance on an acoustically coupled ultrasound transducer limits its clinical applications. To overcome this, PARS [68, 149, 150] was developed, offering non-contact, label-free imaging by replacing the transducer with a co-focused detection laser, as detailed in Section 2.1.4. Recent works have shown that PARS may offer high contrast and high spatial resolution in freshly resected tissue, formalin-fixed tissue, and FFPE tissue sections [17, 150]. In resected tissues, PARS has even been shown to provide 3D imaging of subsurface nuclei [150]. In this thesis, we focus on developing deep learning-driven virtual staining techniques for PARS histology images.

# Chapter 3

# PARS-to-H&E Translation using pix2pix GANs

## 3.1 Introduction

In this chapter, we employ a second-generation PARS architecture [17], to provide label-free histology-like images. As outlined by Ecclestone et al. the PARS architecture incorporates a number of advances, providing improved sensitivity and contrast when imaging in tissue specimens. Applied to resected tissues, PARS directly measures analogous contrast to H&E staining, capturing nuclei and cytoplasm independently. Since the proposed framework directly measures H&E information, it enables confident colorization modeling. In addition, this provides the opportunity for direct comparison to ground truth H&E staining while providing outputs acceptable for clinical applications.

PARS microscopy simultaneously provides label-free, non-contact imaging of radiative absorption, non-radiative absorption, and optical scattering contrasts. To the extent of our knowledge, the PARS is the only modality that can provide all three contrast mechanisms in a single capture. This rich array of input data provides simultaneous recovery of structures including cell nuclei, fibrin, connective tissues, adipocytes, and neurons [17] in a single excitation event. The diagnostic confidence of deep learning-based virtual histological staining can be improved using this more informative dataset for network training and colorization.

Here, we use a general-purpose image-to-image translation model, namely pix2pix [18], to colorize histological images captured by the PARS system. The pix2pix model is designed based on generative adversarial networks (GANs) [15], a type of generative model

that creates new datasets based on a given input (training) dataset. The architecture is comprised of a generator subnetwork for creating new feasible data and a discriminator subnetwork that tries to distinguish between the newly generated (synthetic) data and the ground truth (real) data. The two subnetworks are trained simultaneously in an adversarial process, where the generator model aims to maximize the error rate of the discriminator model, while the discriminator model tries to minimize the classification error. For the training task, the network is fed multi-contrast images of the PARS channels co-registered with H&E images that are captured using traditional slide imaging procedures.

Through the array of contrasts provided by PARS, we directly capture H&E information. Directly measuring the desired H&E contrast in one dataset permits a strong deep-learning model, avoiding assumptions or unsubstantiated inferences about the presence of features such as cell nuclei. The proposed work provides H&E-like PARS of tissue slides, leading to images comparable to the gold standard. PARS virtual histology images were reviewed by four dermatologic pathologists, who confirmed they are of diagnostic quality and that resolution, contrast, and color permit interpretation as if they were H&E. This is an essential step toward developing an alternative histopathological workflow featuring slide-free virtual staining of fresh tissue specimens. Ultimately, achieving histology-like in-situ imaging would permit near real-time intra-operative margin assessment.

## 3.2 Dataset Preparation

The dataset used in this chapter was collected from thin sections of formalin-fixed paraffin-embedded (FFPE) human skin tissues. Anonymous tissue samples were provided by clinical collaborators at the Cross-Cancer Institute (Edmonton, Alberta, Canada). These samples were collected under protocols approved by the Research Ethics Board of Alberta (Protocol ID: HREBA.CC-18-0277) and the University of Waterloo Health Research Ethics Committee (Photoacoustic Remote Sensing (PARS) Microscopy of Surgical Resection, Needle Biopsy, and Pathology Specimens; Protocol ID: 40275). Patient consent was waived by the ethics committee as the samples are archival tissues no longer required for patient diagnostics, and no information was provided to the researchers about the patient's identity. All experiments involving human tissues were conducted in accordance with the government of Canada's guidelines and regulations, such as "Ethical Conduct for Research Involving Humans (TCPS 2)".

The dataset of co-registered PARS and H&E pairs required for the virtual staining procedure is acquired as follows. Unstained, thin tissue sections are imaged with the PARS system. Once imaged, specimens are stained and scanned with a bright-field microscope

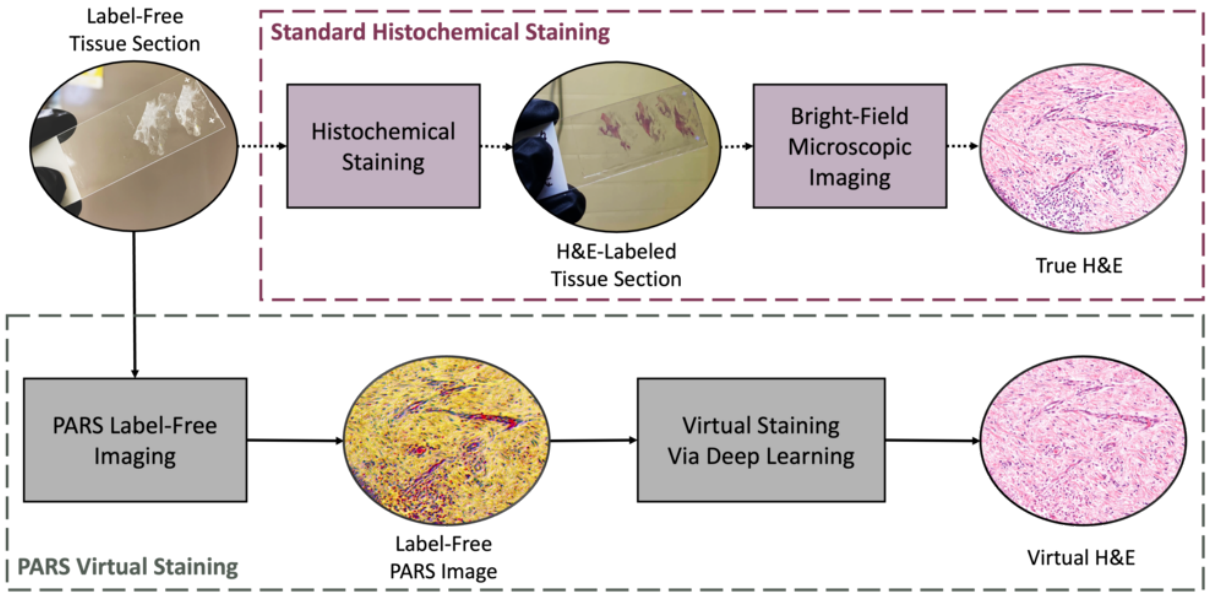



Figure 3.1: Virtual staining of PARS images via deep learning. The top pipeline shows the standard workflow to generate images of histochemical stains. The bottom pipeline shows the steps for virtual staining. The same tissue section is used in the two workflows for deep model training and performance analysis.

to generate matched PARS and H&E image pairs of the exact same samples. The dataset preparation procedure is illustrated in Figure 3.1. In this figure, a comparison is presented between the conventional histochemical staining process and our proposed virtual staining process. It is important to note that, for training a virtual staining model, the samples still need to undergo the standard staining procedure to generate the model's ground truth data. However, this step is not required once the model has been trained.

The unstained tissue sections are prepared as follows. First, bulk resected tissues are fixed in a formalin fixative solution (within 20 minutes of resection) for up to 48 hours. Tissues are then dehydrated and prepared for paraffin wax penetration. During this process, tissues are cleared of residual fats. Next, tissues are embedded into the paraffin substrate, creating FFPE tissue blocks. Thin sections of $\sim 5\ \mu m$ thickness are sliced from the FFPE tissue block surface. Thin sections are fixed directly onto glass slides, and excess paraffin is removed by baking the slide at 60 °C for 60 minutes. At the end of this process, the thin sections are ready for imaging with the PARS system. All imaging is done at the PhotoMedicine Labs at the University of Waterloo. Once the PARS dataset is collected,

the thin tissue slides are stained with H&E, covered with a mounting medium and a cover slip, and then imaged using a transmission mode bright-field microscope, providing the corresponding H&E dataset.

## 3.3 PARS Imaging

A detailed description of the PARS system architecture and image formation process can be found in [17]. Briefly, the experimental system is illustrated in Figure 3.2. The excitation source is a 266 nm diode laser (Wedge XF 266, RPMC). The detection is a 405 nm OBIS-LS laser (OBIS LS 405, Coherent). The excitation and detection sources are combined via a dichroic mirror, then co-focused onto the sample with a 0.42 NA UV objective lens. Detection light returning from the sample (containing the optical scattering and non-radiative relaxation) is fiber-coupled into the circulator, which then redirects the reflected light to a photodiode where the nanosecond scale photoacoustic intensity modulations are captured. Concurrently, radiative relaxation is chromatically isolated and then captured using a photodiode.

The PARS image formation process can be described as follows. Mechanical stages are used to scan the sample while the objective lens remains fixed. The velocity of the stages is adjusted to obtain a pixel size of 250 nm. The 50 kHz pulsed excitation source is used to excite the tissue, and the induced signals due to optical scattering, radiative relaxation, and non-radiative relaxation are collected at each pulse. Each event is placed in the channel that corresponds to its signal. Such signals are compressed by extracting a single feature, which becomes the pixel value. An image is then reconstructed for each PARS channel by fitting the computed pixel values to a Cartesian grid based on their corresponding position signals from the stages. The images are then processed according to the task of interest. In the next subsection, the processing of images for the virtual staining task is described.

## 3.4 Dataset Preparation for Training

In this study, we explore the utilization of the pix2pix GAN model [18] for virtual staining. This choice stems from the model's extensive usage, as discussed in Section 2.2 and the advantageous availability of paired data, making it a suitable candidate for our investigation. The model requires pixel-to-pixel correspondences between the input data (PARS) and the ground truth (H&E), where the objective is to learn a statistical transformation between the two data domains. Since PARS and H&E images are acquired using different

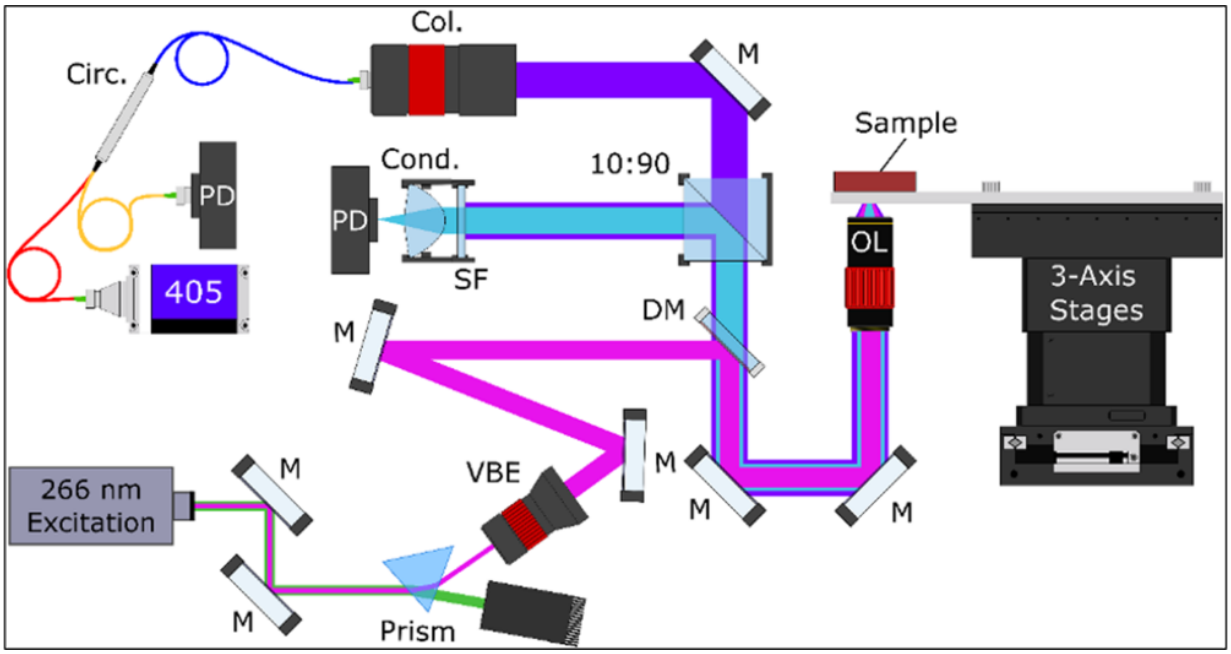



Figure 3.2: Simplified PARS system architecture and system setup. Component labels are as follows: dichroic mirror (DM), variable beam expander (VBE), collimator (Col.), circulator (Circ.), spectral filter (SF), condenser lens (Cond.), photodiode (PD), 10:90 splitter (90:10), mirror (M) and objective lens (OL). Source: Reprinted from [1].

techniques, images of the same sample are not co-registered. Therefore, pre-processing, including the field of view matching and registration, is required to produce approximately co-registered pairs of PARS and H&E images. The virtual staining framework is depicted in Figure 3.3. The registration process can be described as follows. First, the corresponding field-of-views (FOVs) of the two image domains are extracted from the whole-slide images and coarsely matched in terms of pixel size in preparation for the one-to-one registration process. For registration, the control point registration tool is deployed from the MATLAB Image Processing [151] and Computer Vision [152] Toolboxes. The PARS images are chosen as reference images, and H&E as the images to be registered. The control points are manually selected and then fine-tuned in a second pass to minimize global and local distortions. Afterward, the algorithm fits a non-rigid geometric transformation [153] between the two images, which is applied to the H&E images to obtain co-registered PARS and H&E images. Since all PARS channels are intrinsically registered, the whole registration process is applied only to the non-radiative channel.

Once the registration process is done, the dataset is ready for network training for colorization. Here, all three PARS channels are used for training the model. Figure 3.4 illustrates the inputs and the output of the network. Non-radiative and radiative absorption and optical scattering channels are shown in Figure 3.4(a-c), respectively. The multi-channel input is shown in Figure 3.4(d) as the concatenation of the three channels. A colorization example is depicted in Figure 3.4(e) and the corresponding H&E ground truth is shown in Figure 3.4(f). It is clear that the non-radiative absorption channel features strong nuclear contrast representing the hematoxylin contrast, while the radiative absorption contrast exhibits the eosin contrast showing the tissue morphological information. Normalized PARS images undergo contrast stretching by saturating the top 1% and the bottom 1% of all the pixel values and are then color-reversed to match the colormap and the histogram distribution of the grayscale ground truth images. Several datasets are prepared in a likewise manner for concept validation. This framework is meant to replace the histological staining process, where the pix2pix model is trained using the multi-channel PARS images as inputs and corresponding H&E images as labels, and the network is trained to learn the statistical transformation function between the two domains. Once the model is trained, the virtual staining process takes a few seconds to produce a virtually stained version of an image of 4000 x 6500 pixels ($\sim$1 $mm^2$).

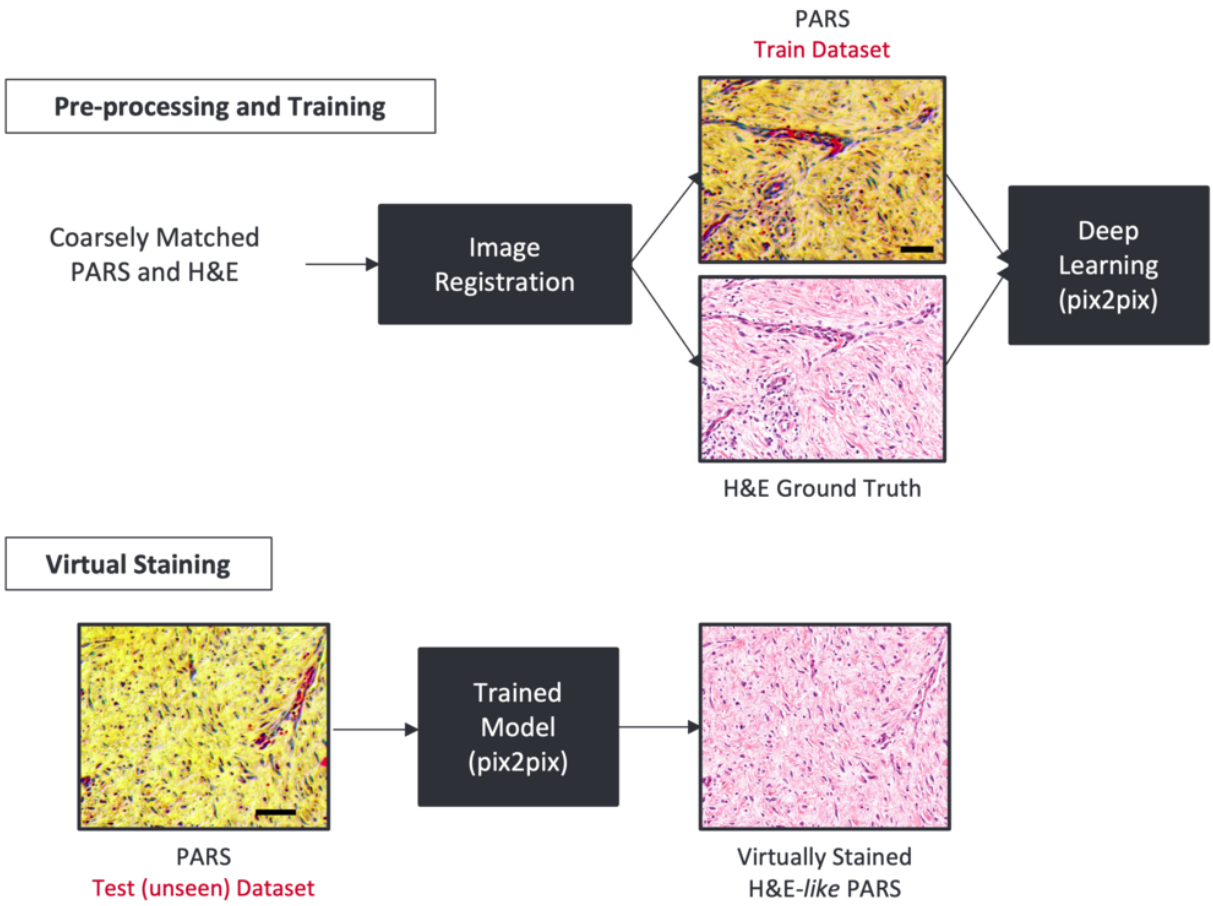



Figure 3.3: Virtual staining framework. Pairs of PARS and H&E images are co-registered to create the training dataset, which is then used to train the pix2pix [18] model. Unseen PARS images are then fed into the trained model for virtual staining, producing H&E-like visualizations. Scale bar: 50 $mum$.

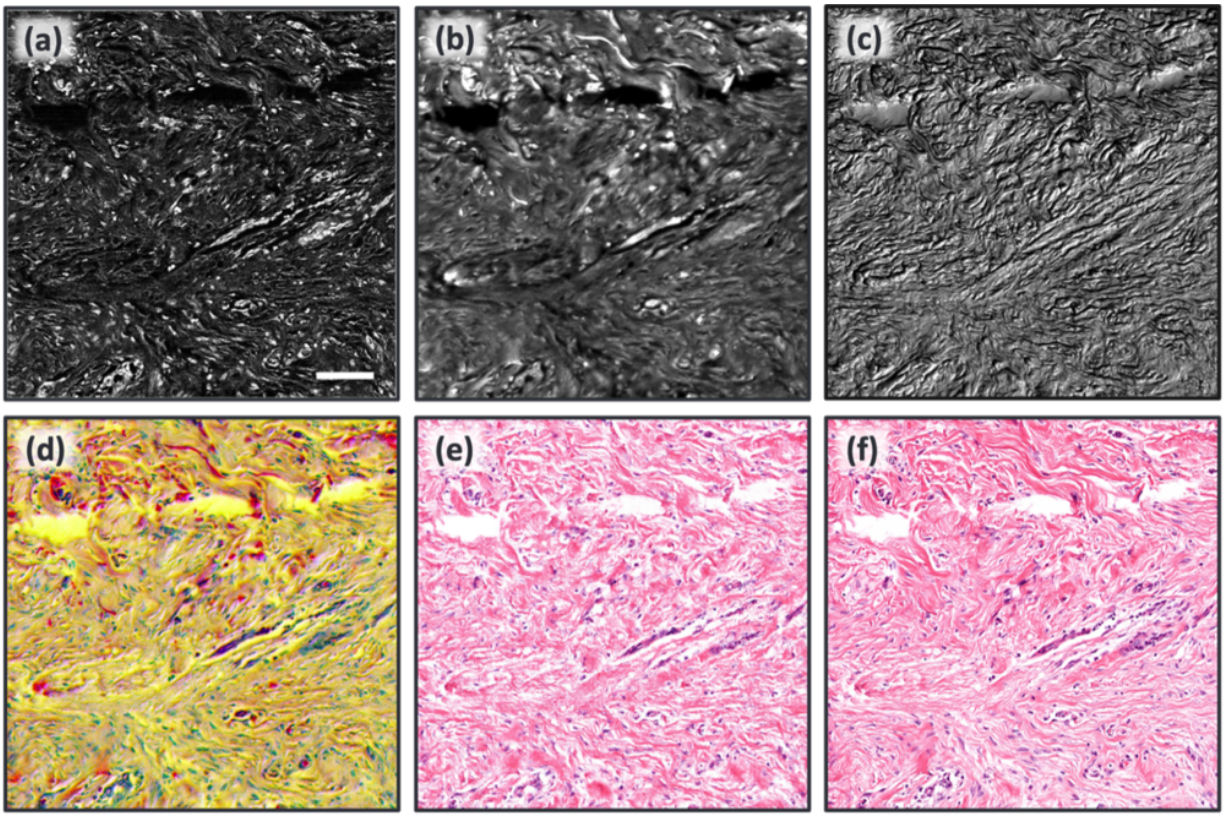



Figure 3.4: The colorization network input and output images of human skin tissue slide. (a) PARS non-radiative absorption contrast image captured using a PARS microscope. (b) PARS radiative absorption contrast of the same tissue using the same microscope. (c) PARS optical scattering contrast. (d) PARS image of all three channels (network input to be colorized). (e) Colorized PARS image using pix2pix model. (f) Histologically stained tissue of the same region taken on a bright field microscope, the ground truth for the virtual staining framework. Scale bar: 100 $\mu m$.

## 3.5 Implementation Details

A pix2pix image-to-image translation model [18] is applied to transform the label-free PARS images to virtually stained images that match their corresponding H&E histochemical stained samples. The model is based on a GAN architecture, or more specifically conditional GAN (cGAN) [103], where the generative model is further conditioned on the input data.

The model training algorithm in pix2pix concurrently trains two separate neural networks, the generator, and the discriminator, as illustrated in Figure 3.5. The training algorithm learns to minimize a loss function between the network output and the ground truth. The discriminator network is trained to effectively differentiate between real and synthetically generated images, while the generator network is trained at the same time to produce images that better resemble the ground truth data, which makes it progressively harder for the discriminator to differentiate correctly. As a result, the model is capable of producing virtually stained PARS images of comparable diagnostic quality with the histologically stained images.

Training for the standard pix2pix model is carried out using 3-channel RGB color images. In the proposed approach, the three channels are replaced with non-radiative absorption contrast, radiative absorption contrast, and optical scattering, respectively. In this way, the model can simultaneously learn complementary information from the three available information sources.

Overlapping patches of 256×256 pixels are extracted from PARS and H&E images to generate the training and validation sets. Approximately, 15000 training patches and 6000 validation patches are used in the experiments. The maximum number of epochs is set to 500, with an early stopping criterion to terminate the training when the generator loss stops improving. The trained model is then applied to the test images, which are also subdivided into overlapping patches of 256×256 pixels. An overlap of $\sim 50\%$ is usually sufficient to avoid visible artifacts at the borders of adjacent patches in the final stitched image. The colorization algorithm is implemented in Python version 3.8.10, and model training is implemented using PyTorch version 1.9.1 with support for CUDA version 11.

## 3.6 Clinical Validation of Virtual H&E Images of Skin

Nine representative images of PARS-generated, and AI-colorized human skin were distributed to four board-certified dermatopathologists for evaluation. None of the der-

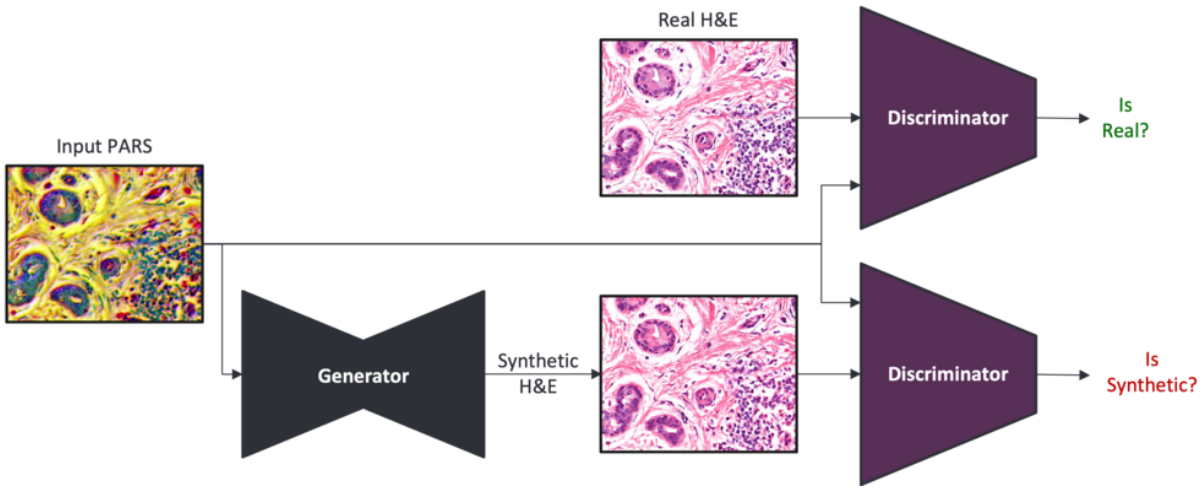



Figure 3.5: An illustration of the pix2pix training algorithm. Pix2pix is comprised of two subnetworks: the discriminator and the generator. The discriminator subnetwork is trained to distinguish between real H&E and synthetically generated H&E. Simultaneously, the generator subnetwork is trained to confuse the discriminator by producing images that look like real H&E.

matopathologists has conflicts of interest or financial relationships relating to this technology. Each pathologist confirmed that all the images are of sufficient quality to allow diagnosis, and that contrast, resolution, and colorization quality are sufficiently H&E-like to permit clinical diagnoses.

## 3.7 Results and Discussion

We applied our machine learning framework to unstained human skin tissue sections imaged with PARS. These results are compared against bright-field microscopy images, which were acquired after staining the exact same tissues with H&E. Hence, these images provide a one-to-one nuclei-to-nuclei match between the PARS and corresponding H&E images. Given limited datasets, our initial step is to perform a proof of concept and bracket the model's performance by overfitting. When overfitting, the same data is used for training and testing the model. This is performed using a single image, split into 5000 patches of 256×256 pixels each. The colorization results of this experiment are visually almost indistinguishable from the ground truth images (as shown in Figure 3.6), setting an upper bound on the model's performance. When generalizing the model, comparable results can

be realistically achieved with data not included in the training dataset (unseen data) if the training set sufficiently covers the variability of the target domain.

With the initial conceptualization of the colorization process examined through the overfitted data processing, the next step is a generalization. Hence, we expanded our experiments to assess the model's generalization performance. Approximately, 15000 training patches and 6000 validation patches are used in the experiment. The model is tested exclusively on unseen PARS images, which are taken from the same label-free tissue sections. Again, the results show strong visual agreement between the PARS virtual staining and the H&E ground truth, where the colorization of different tissue structures resembles what is usually seen in H&E-stained tissue samples. Specifically, of interest in the tissues shown in Figure 3.7, are the nuclei. In many malignancies, the morphology and distribution of individual nuclei may be indicative of the progression and severity of the disease. Hence, accurately representing the nuclear structures is paramount for any virtual staining technique. As demonstrated in Figure 3.7 there is high acuity in the replication of the nuclei's shape and size between the colorized PARS and the true H&E. Assessing the colorized section in Figure 3.7(a), the resemblance is even more apparent as the nuclear structures are highly similar between the H&E and the virtual staining.

In addition to the visual agreement, two statistical metrics are used for comparison, the structural similarity index measure (SSIM) [126] and the root mean square error (RMSE). The colorized and ground truth images are converted to the Lab color space before SSIM and RMSE computations to represent the images in a perceptually correlated color space. The RMSE is chosen to provide a direct comparison between the images, capturing true differences between the virtual stain and the H&E. The SSIM is selected as a metric since it measures the perceived change in structural information as well as the contrast and luminance changes, rather than the absolute errors, mimicking the human visual system [126]. This metric (SSIM) is of particular interest due to the subjective evaluation nature of histopathology [154]. The statistical metrics are applied on 1000 patches of size $256 \times 256$ pixels. The average computed SSIM is $0.91 \pm 0.02$ while an average RMSE of $14.28 \pm 1.61$ is measured, indicating objective structural and quantitative agreement between the H&E-like PARS and true H&E.

Assessing additional, larger skin tissue sections (Figure 3.8), a further comparison may be drawn between the diagnostic features of the PARS and the H&E images. These images capture the dermis of human skin tissue containing predominately connective tissue and blood vessels. Directly observing the raw PARS data (Figure 3.8-i) the nuclear structures are discernable in the dark blue/purple, while the connective tissues are seen in shades of yellow and orange. Finally, the microvasculature is distinctly visible in red. Specific details such as the vasculature and nuclei may be easier to identify and assess in the raw

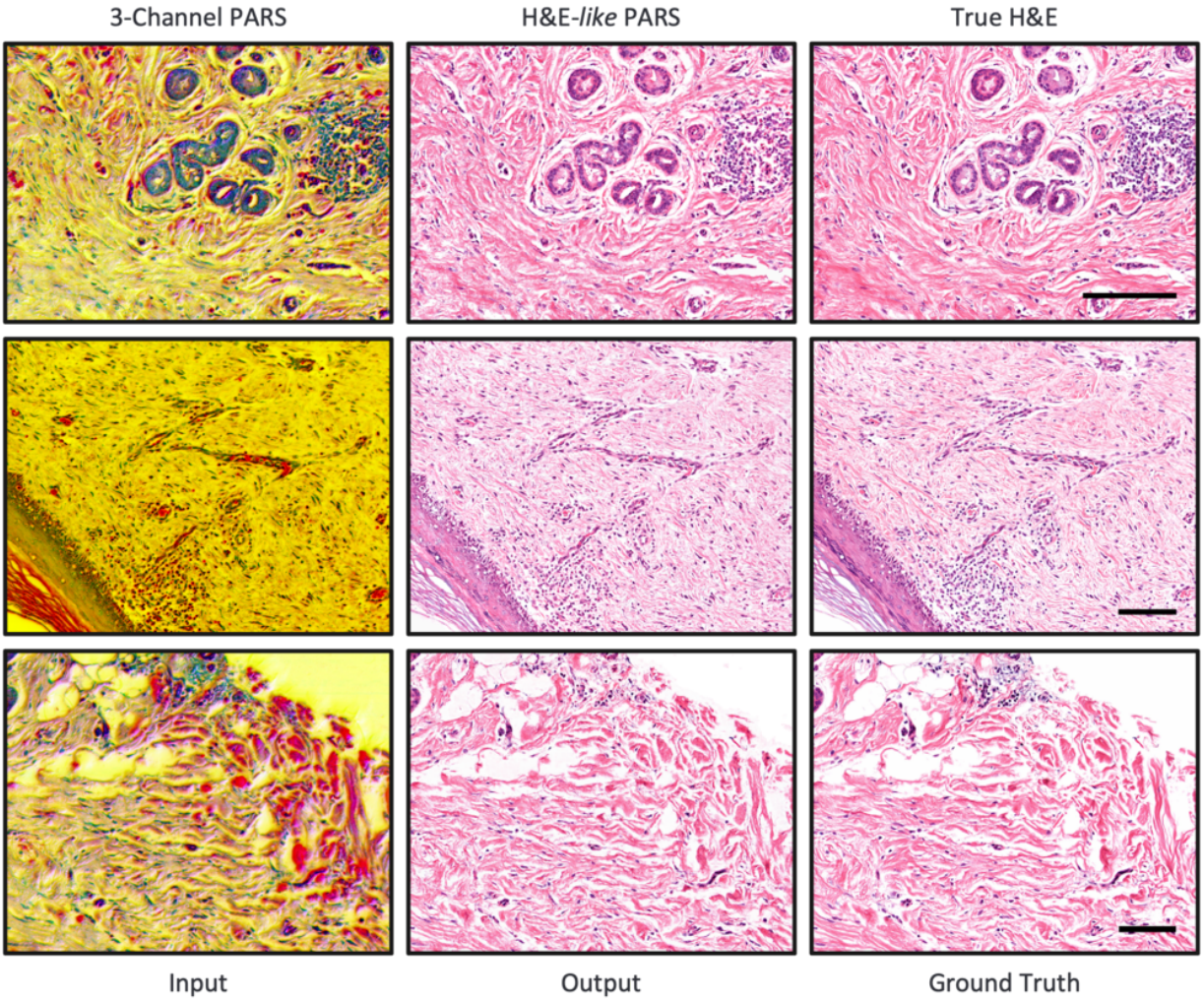



Figure 3.6: Virtual staining of PARS images of human skin tissue slides. Here, the same images are used for training and testing the model, setting an upper bound on the model's performance. Scale bar: 100 $\mu m$.

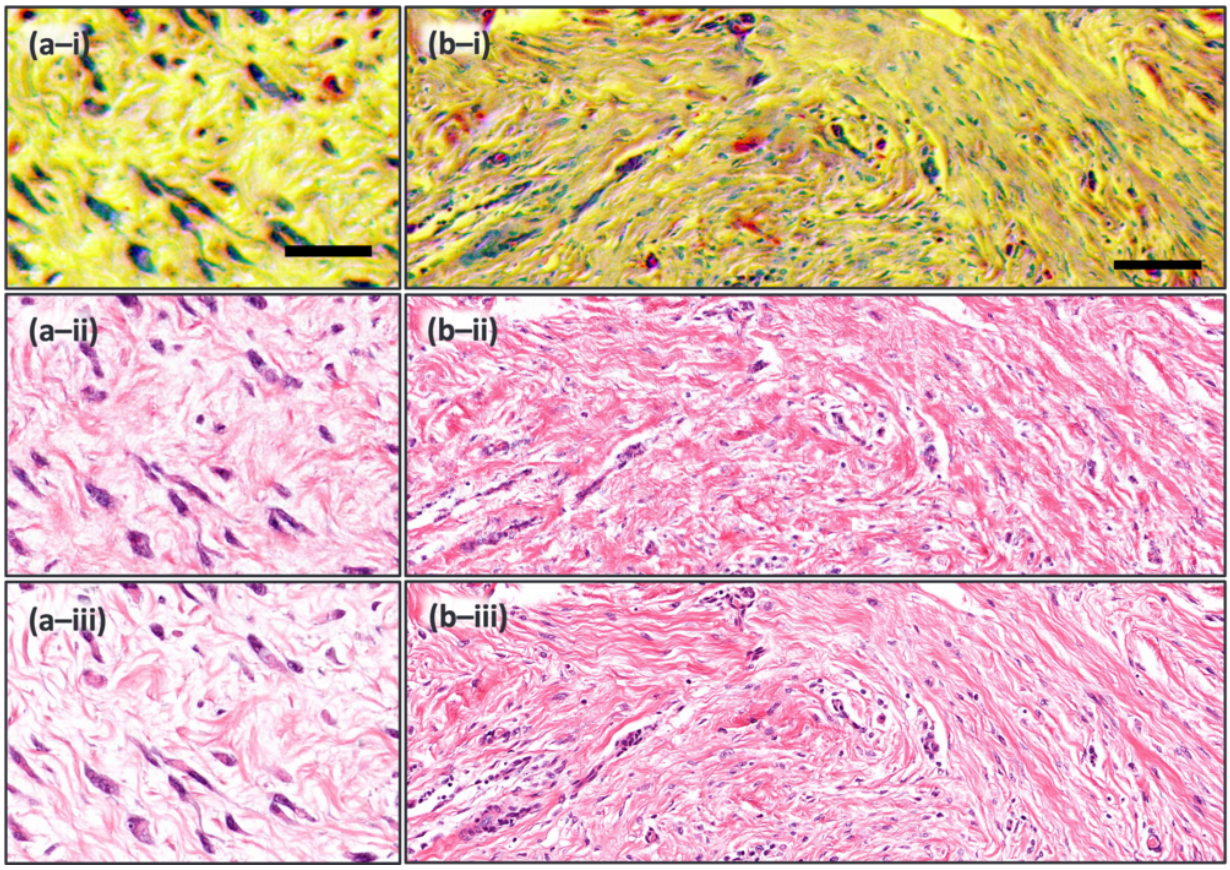



Figure 3.7: Virtual staining results of different parts of human skin tissue using unseen data. (a) Magnified image of connective tissue. Scale bar 25 $\mu m$. (b) Image of dense irregular connective tissue. Scale bar: 100 $\mu m$. (i) Raw PARS. (ii) H&E-like PARS. (iii) True H&E. Image sets (a) and (b) depict high visual agreement, especially of nuclear structures, between the H&E and the virtual staining.

PARS input (Figure 3.8-i) as compared to the H&E (Figure 3.8-iii) or the virtually stained images (Figure 3.8-ii). All the diagnostic information required for a pathologist to perform diagnosis and histological analysis may be present in the raw PARS images. However, the PARS manifestations are colorized significantly differently than the gold standard H&E images. Current pathologists have extensive training and pattern recognition skills based primarily on H&E imaging. Hence, substantial retraining would be required for pathologists to interpret the different colors in the raw PARS images. This further motivates the proposed AI colorization technique, which may enable clinical acceptance by transforming the PARS images into a clinically accepted format.

With PARS virtual H&E, both the collagen-rich subcutaneous connective tissues and nuclear structures are represented with high quality and contrast, effectively comparable to the gold standard imaging. Furthermore, four independent dermatopathologists evaluated the images and confirmed that the colorization, contrast, and resolution are sufficient for clinical diagnosis and subjectively equivalent to H&E imaging. This is an encouraging conclusion as these PARS images are produced directly from unstained tissue specimens. Though the current work is restricted to tissue sections to have a one-to-one correspondence with H&E from the same tissue, PARS microscopy also provides high-resolution images of freshly resected tissues [17], directly capturing hematoxylin and eosin contrasts enabling comparable fresh tissue colorization.

Finally, an experiment is carried out to verify the contribution of the PARS optical absorption contrasts to the virtual staining process. Three different models are trained: one using the non-radiative channel alone, one using the radiative channel alone, and one using all three contrasts. The models are trained using the same training settings and the same training and validation sets described earlier in this section. Looking at the results in Figure 3.9, it is evident that using multi-channel input yields better results than using single channels. For instance, some of the nuclei are confused by the model as connective tissue when using the radiative channel only, leading to pink colorization of nuclei instead of purple. On the other hand, using only the non-radiative channel results in fewer pink shades of the different connective tissue structures as shown in the figure.

It is important to highlight that pix2pix, being an image-to-image style transfer model, demonstrates the ability to learn not only the color but also the resolution characteristics of the target domain, a technique known as super-resolution [99, 155]. This observation is consistent with our findings from numerous experiments. As evident in Figure 3.9, when using only the radiative channel, which exhibits a noticeably lower resolution than the target domain, the colorized results maintain a high resolution, closely resembling the ground truth data. This indicates that the model successfully incorporates both color and resolution information during the learning process for virtual staining.

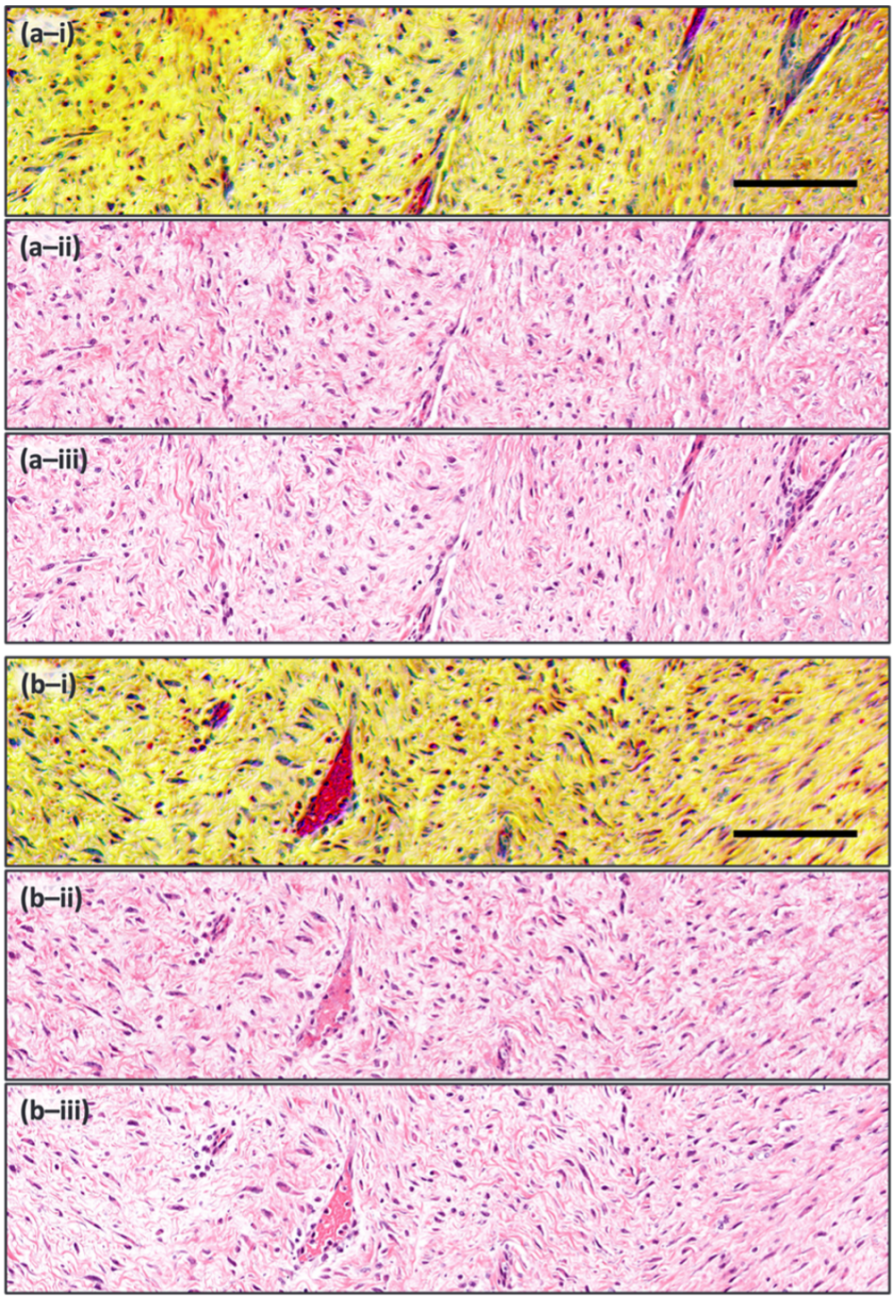



Figure 3.8: Another set of virtual staining results using unseen data of human skin tissue. Image sets (a) and (b) show different parts of the dermis like connective tissue and blood vessels. (i) Raw PARS. (ii) H&E-like PARS. (iii) True H&E. Scale bar 100 $\mu m$.

SSIM and RMSE are also applied on 1000 patches of size 256 x 256 pixels. When using a single-channel input, the average computed SSIM is 0.88 ± 0.02 and the average RMSE is 16.39 ± 1.25 for the non-radiative channel, and an SSIM of 0.89 ± 0.02 and RMSE of 15.87 ± 1.48 for the radiative channel. These statistical metrics objectively prove that multi-channel inputs significantly improve the virtual staining results. The results also reveal the contribution of each channel, where the non-radiative channel mostly contributes to the colorization of the nuclear structures, while the radiative channel assists with the pink shades. In conclusion, providing the model with multi-modal data instead of single-mode inputs greatly improves the overall virtual staining quality.

In the near future, formalized clinical studies will be conducted aiming to validate and compare the PARS virtual H&E against traditional visualizations. Once this validation is complete proving the accuracy of the colorization, the proposed system may be employed directly in bulk resected tissue specimens. When moving to bulk tissues, PARS imaging provides other potential advantages over standard H&E. In addition to the markedly reduced time to diagnosis, this technique, when applied to fresh tissue, removes numerous pre-analytic variables such as the time since devitalization, time from devitalization to fixation, time of fixation, variability of fixation depending on the distance from the surface, and staining variability. This could improve the consistency of histology evaluation by removing operator-dependent variability in tissue processing.

## 3.8 Conclusion

In summary, a framework is developed for H&E-like virtual staining of label-free PARS images. The proposed method leverages the information directly captured from all three PARS channels (radiative absorption, non-radiative absorption, and scattering), providing a robust model that successfully generates H&E-like visualizations of tissue samples. Feeding the model with the multi-modal data provides a plethora of data not afforded by any single mode. This may markedly improve the overall colorization quality. Specifically, this method aims to avoid inferring tissue features, to achieve diagnostically accurate colorizations. Applied in thin sections of human skin tissues, the proposed method achieves a high degree of agreement between virtual staining and gold-standard histological staining. Dermatopathologists reviewing the PARS colorized images unanimously confirmed the diagnostic adequacy of these results. Four independent pathologists indicated that the colorization, resolution, and contrast are acceptable as a H&E surrogate. Applied to clinical settings, this may allow for unstained tissues to be imaged, directly producing H&E-like visualizations. Such a technique could reduce histological imaging times while

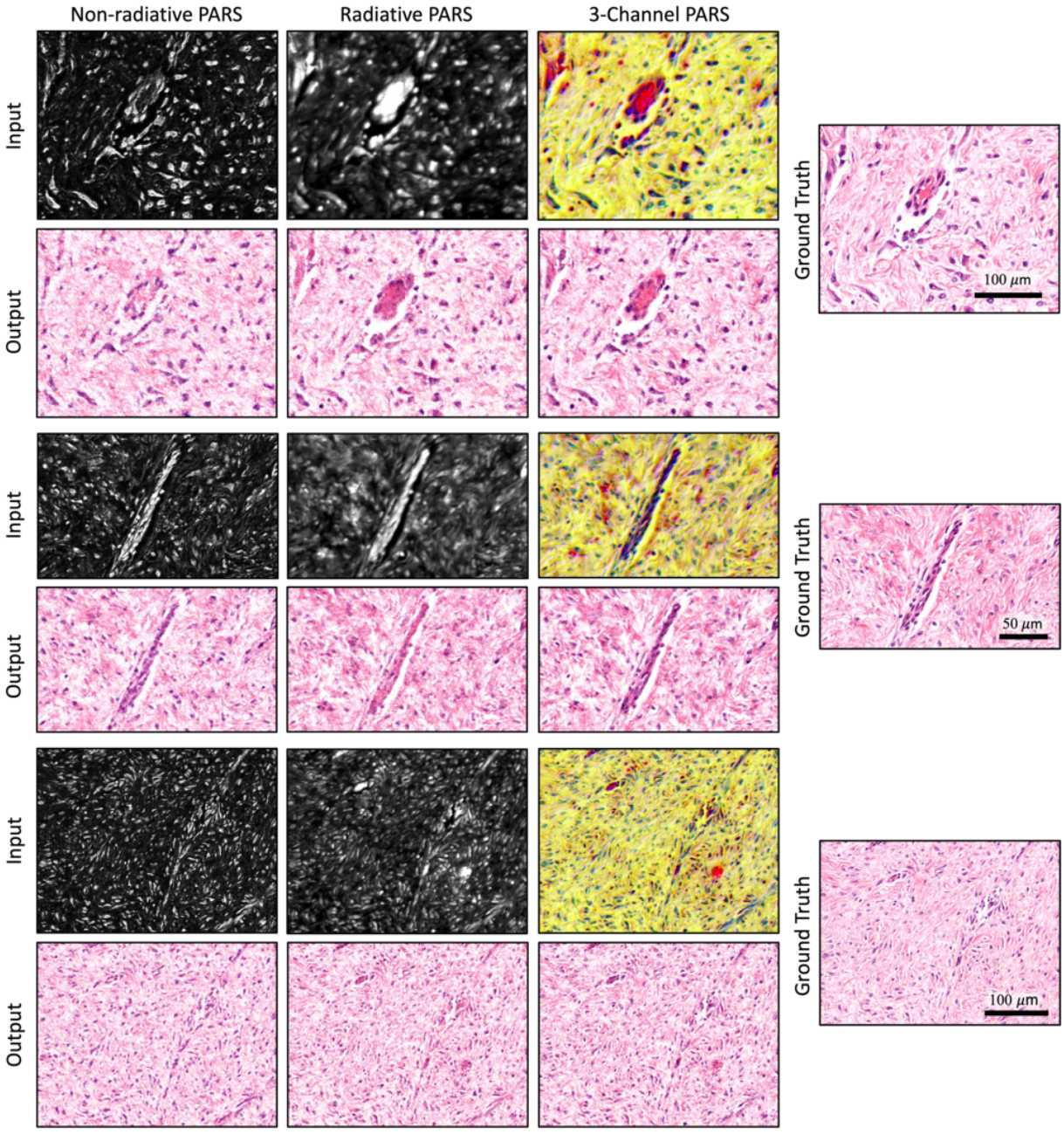



Figure 3.9: A comparison between virtual staining outputs using different inputs: non-radiative PARS channel only, radiative channel PARS only, and all three PARS channels.

simultaneously reducing staining variability by removing operator dependencies within the histochemical staining process.

# Chapter 4

# A Comparison between Pix2pix GAN and CycleGAN for PARS-to-H&E Translation

## 4.1 Introduction

PARS microscopy has proven its effectiveness in producing digital representations of H&E stains through the utilization of the pix2pix model [1], as presented in Chapter 3. However, the successful application of pix2pix requires precise alignment of PARS and true H&E images at the individual pixel level, which can be challenging and often results in colorization that is distorted or blurred due to registration inaccuracies. The challenges related to image alignment mainly stem from the fact that the PARS data and the corresponding true H&E images are acquired using different imaging techniques. Moreover, the presence of a complete staining process between the two image captures can introduce deformations within the tissue specimen, which further compounds the challenges of achieving accurate image registration. Registration errors can significantly impact the quality of colorization results. These errors manifest as smudged appearances and noticeable blurring of the majority of structures, especially in the case of highly detailed images. This is the case even when the structures are well-defined in the original raw data.

In contrast, an unpaired image-to-image translation model called cycleGAN [19], short for cycle-consistent generative adversarial network, was specifically designed to address the limitations posed by paired data. CycleGANs extend the capabilities of the pix2pix architecture by enabling the model to handle unpaired image collections. This is achieved

through the introduction of a cycle-consistency loss term, which allows the model to establish meaningful relationships between two different image domains. This loss term ensures that when an image is translated from one domain to another and then back, it returns to its original form. By enforcing this constraint, cycleGANs learn the mapping between the domains, facilitating image translation and style transfer without paired examples. CycleGANs have significantly advanced the field of image-to-image translation and provided a flexible framework for exploring diverse image transformations.

As highlighted in Table 2.1, cycleGAN and its variations are frequently employed in unsupervised training scenarios when there is no availability of paired data [122]. This preference for cycleGAN can be attributed to its ability to generate images that retain the underlying content while adopting a style similar to the training data [156]. Conversely, methods such as DualGAN [157] may not effectively preserve the content of the image [156]. In the case of other image-to-image transformation models like CUT [158], the content remains intact, but style transfer may introduce color artifacts into the transformed image [156]. Among unpaired image-to-image translation models, cycleGAN outperforms others, as evidenced by superior scores in structural similarity and chromatic accuracy [159]. This achievement culminates in a color distribution closely resembling chemical staining. Consequently, in tasks demanding the strict preservation of both content and style, as is often the case in virtual staining, the cycleGAN's cycle consistency can be advantageous.

While cycleGANs have shown promise in various image-to-image translation tasks [160, 161, 162, 163], there are specific downsides when considering their application in virtual staining and diagnosis. One significant drawback arises as a result of the lack of direct ground truth during the training process, as unpaired data is used. This can lead to inaccuracies in the generated virtual stains, potentially affecting the reliability of the diagnostic outcomes. These limitations highlight the importance of continued research and development to refine cycleGANs to ensure their suitability and reliability in virtual staining and diagnosis applications. This Chapter involves direct comparisons between the previously introduced pix2pix model from Chapter 3 and cycleGAN in the context of virtual staining.

## 4.2 Methods

### 4.2.1 Sample Preparation

The tissue samples used in this chapter underwent a standardized preparation process. Firstly, the samples were fixed in a formalin solution within 20 minutes of excision and left

for 24 to 48 hours. Then, they were dehydrated using ethanol and treated with xylene to remove residual ethanol and fats. The samples were embedded in paraffin wax, resulting in formalin-fixed paraffin-embedded (FFPE) blocks. Thin tissue sections (approximately 4-5 $\mu$m) were cut from these blocks using a microtome. These sections were placed on glass microscope slides and subjected to baking at 60°C for around 60 minutes to eliminate excess paraffin.

Initially, the unstained samples were imaged using the PARS microscope at 40x magnification, followed by staining with H&E. The H&E-stained slides were then imaged at 40x. This entire process was carried out on various malignant human skin samples, from which whole slide images were obtained. The tissue samples were provided by clinical collaborators at the Cross-Cancer Institute in Edmonton, Alberta, Canada, and all patient identification was removed from the samples. Ethical consent was waived since these archival tissues were no longer required for patient diagnostics. The experiments were conducted following approved protocols by the Research Ethics Board of Alberta (Protocol ID: HREBA.CC-180277) and the University of Waterloo Health Research Ethics Committee (Photoacoustic Remote Sensing (PARS) Microscopy of Surgical Resection, Needle Biopsy, and Pathology Specimens; Protocol ID: 40275), in accordance with Canadian government guidelines and regulations which includes "Ethical Conduct for Research Involving Humans (TCPS 2)".

### 4.2.2 Sample Imaging and Image Formation

Here, label-free images were obtained using the whole-slide scanning PARS system, previously described by Tweel et al. [164]. The system employs a 400 ps pulsed UV laser with a frequency of 50 kHz and a wavelength of 266 nm (Wedge XF 266, RPMC) to excite the sample, inducing both non-radiative and radiative relaxation processes. Time-resolved signals resulting from non-radiative relaxation are probed using a continuous-wave 405 nm detection beam (OBIS-LS405, Coherent). To focus the detection beam and the excitation light onto the sample, a UV objective lens with a numerical aperture (NA) of 0.42 is employed. The radiative emissions and the transmitted detection light are gathered independently by another objective lens of NA 0.7 (278-806-3, Mitutoyo). These collected light signals are then directed to an avalanche photodiode (APD130A2, Thorlabs) for detection.

To form an image, mechanical stages move the sample in an "s"-like scanning pattern to laterally separate the excitation events on the sample ($\sim$250 nm/pixel). At each excitation pulse, several hundred nanoseconds of time-resolved signal from each system photodiode are digitized at a 200 MHz rate (CSE1442, RZE-004200, Gage Applied). A portion of the

collected signal is pre-excitation and is used to form the scattering image of the sample in its unperturbed state. The non-radiative image pixels are then derived as a percentage modulation in the detection scattering (post-excitation). Next, the radiative image pixels are obtained from the peak emission amplitude recorded after each excitation event. Pixels are then arranged in a cartesian grid based on the stage position feedback, forming a stack of three co-registered label-free image contrasts: non-radiative, radiative, and scattering. Finally, the excitation pulse energy and detection power, recorded throughout imaging, are used to correct image noise caused by laser power and pulse energy variability.

Whole slide samples can be scanned using the automated workflow previously described by Tweel et al. [164]. In brief, the entire tissue area is divided into subsections (500×500 $\mu$m), each individually scanned at its optimal focus position. Using their relative stage positions and a small amount of overlap ($\sim$5%), these sections are stitched and blended into a single whole slide image.

### 4.2.3 Image Registration

In the previous chapter, we employed a specific approach for image registration. This process involves aligning corresponding features or points in two different images to ensure proper spatial alignment. In the work presented in this chapter, we apply this image registration technique to various pairs of images. Specifically, we conduct both loose and tight registrations (explained below) to assess their impact on the results of pix2pix colorization. Through this evaluation, we aim to understand how image registration influences the accuracy and quality of colorization in different scenarios.

Unlike pix2pix, cycleGAN's performance is less sensitive to the degree of registration [19]. In fact, we observe that loose registration suffices for cycleGAN, meaning that aligning corresponding features in a less stringent manner still produces satisfactory colorization results as presented in Section 4.3.

In our image registration process, we adopt the projective geometric transformation method [165], which has proven to be effective for aligning images in Chapter 3. This transformation requires a minimum of four control point pairs to compute the necessary geometric adjustments. We designate the approach using only four points as "loose registration," as it provides a basic alignment but may not be sufficient for some complex cases where images have significant variations or deformations.

To achieve more accurate and precise registration for the specific dataset at hand, we found that a greater number of control points is essential. Through our experiments, we have noticed that having at least twelve control points for each PARS image noticeably

enhances the registration quality. The use of twelve points is approximately equivalent to having 1 to 2 control points per 1000×1000 pixels in the PARS images. We refer to this enhanced registration strategy as "tight registration." An illustration of the MATLAB GUI [151, 152] employed for registration along with the control points manually assigned for both loose and tight registration is shown in Figure 4.1.

Figure 4.2 (a) and (b) illustrate the outcomes of both the loose and tight registration processes, respectively. In the overlay of PARS and H&E images depicted in (a), using only four points results in some regions of acceptable registration, while others show noticeable shifts. This observation suggests the occurrence of local deformations, indicating the need for more points to achieve a better alignment. On the other hand, the overlay presented in (b) demonstrates a significant improvement in alignment when utilizing 12 points during the registration process.

It is important to highlight that the PARS images depicted in Figure 4.1 are derived from the PARS non-radiative channel. To facilitate the registration process, we take advantage of the inherent registration among all PARS channels. Consequently, rather than utilizing multiple channels, we opt to use just one channel from the PARS images specifically for the registration procedure. We choose non-radiative images for registration because the nuclei in these images possess unique characteristics that make them ideal control points. Their small, round shape provides distinct features for precise alignment in both PARS and H&E images, using the nucleus center as a reliable reference. Additionally, their high contrast in both image domains allows for accurate localization during the registration process. As a result, nuclei serve as robust fiducial markers, ensuring effective alignment between PARS and H&E images.

### 4.2.4 Implementation Details

The pix2pix model employed here was identical to the one used in Chapter 3. As for cycleGAN, we utilized the implementation described in [19], which is detailed in Section 2.1.6.

Two datasets were created for pix2pix training, comprising loosely and tightly registered pairs. Regarding cycleGAN, only the loosely registered dataset is utilized for training. To generate the training and validation sets, overlapping patches of 256×256 pixels were extracted from both PARS and H&E images. The experiments involved using approximately 4500 patches, with a 70% split for training, 10% split for validation, and 20% for testing.

The training regimen for the two models was established with certain specifications: a maximum of 200 epochs were allowed, accompanied by an early stopping criterion that

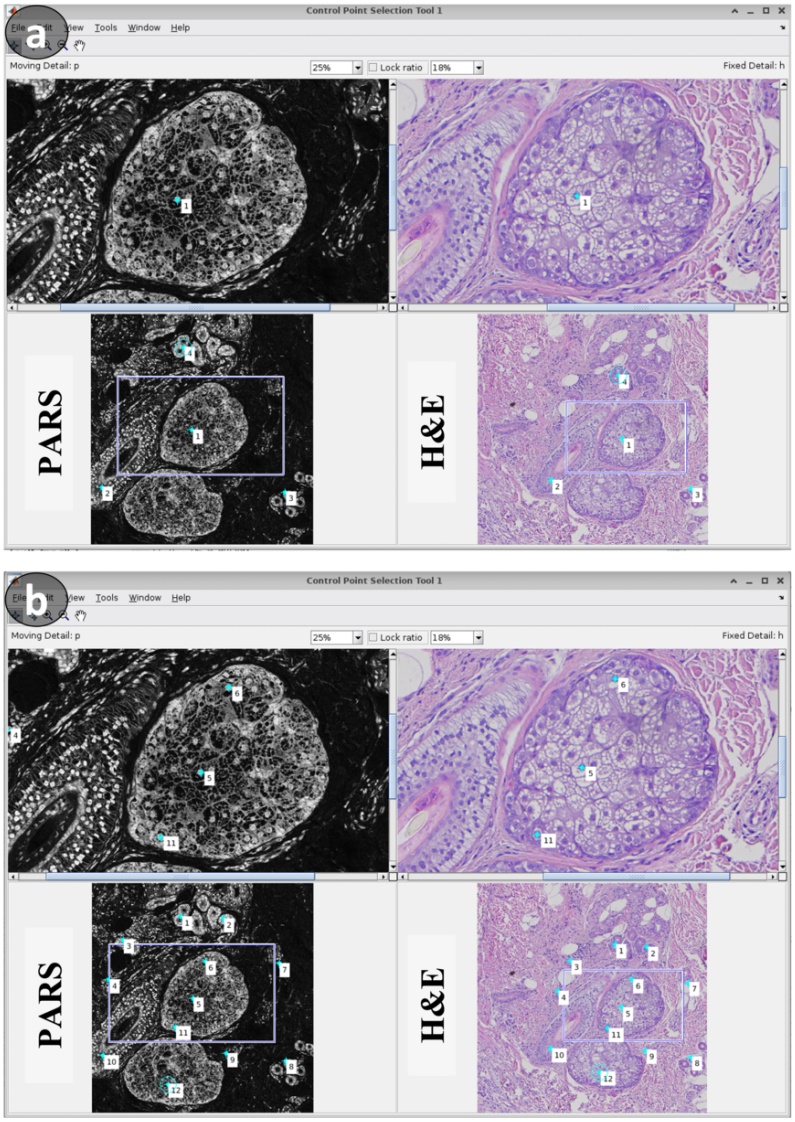



Figure 4.1: MATLAB GUI [151, 152] used for registration of PARS and H&E images. (a) Loose image registration using 4 control points. (b) Tight image registration using 12 control points.

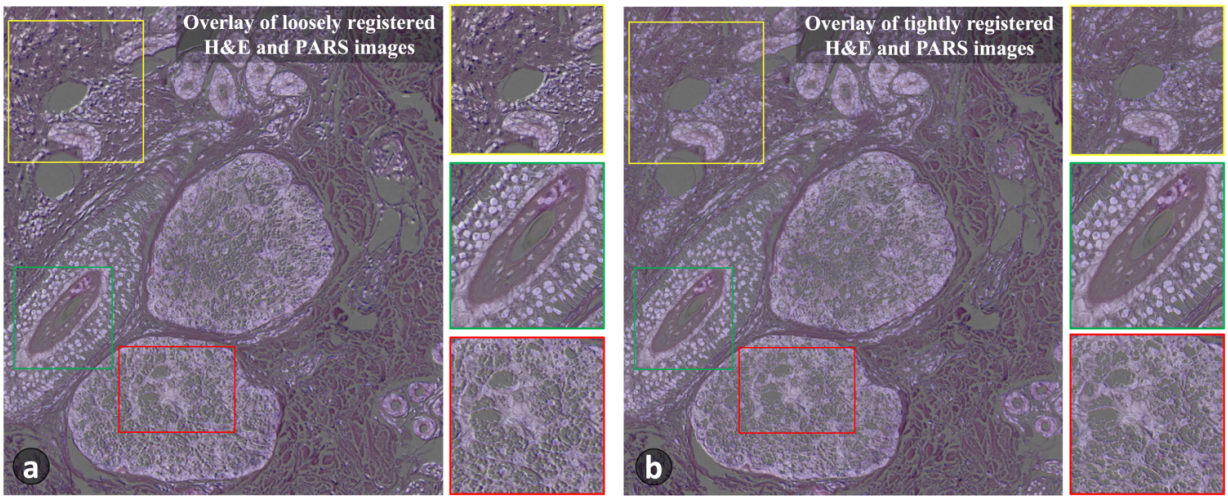



Figure 4.2: The overlay displays the image registration results of PARS and H&E images. (a) Loose registration results using 4 control points. (b) Tight registration results using 12 control points.

ceased training when the generator loss plateaued. The choice of the number of epochs was based on empirical observations and adhered to the recommended value outlined in the original pix2pix and cycleGAN implementations [18, 19]. Additionally, a learning rate of 0.0002 was employed, and it was subjected to linear decay every 10 epochs until the training completion. This specific learning rate was identified as the most effective through experimentation with different values, aligning with the recommendation provided by the authors [18, 19]. All other parameters including the Adam solver for the optimization algorithm and the activation functions were kept the same as the original frameworks.

Following training, the models were applied to test images taken from another tissue slide, which were also divided into overlapping patches of 256×256 pixels. An overlap of 80% was determined to be effective in preventing any noticeable artifacts at the boundaries of neighboring patches in the stitched final image. This ensured a seamless and coherent representation of the test images after colorization using the trained models.

## 4.3 Results and Discussion

The virtual staining models were applied to compare the performance using unstained human skin tissue sections captured by the PARS microscope. To validate our results, we

compared them against corresponding bright-field microscopy images of the same tissues, which were obtained after staining with H&E. The comparison allows for assessing the accuracy and fidelity of the colorization process for different models while also providing a baseline reference.

For our experiments, three distinct models were trained: pix2pix using the loosely registered dataset, pix2pix with the tightly registered dataset, and cycleGAN using the loosely registered data. To understand the effect of image registration on pix2pix colorization results, we conducted both loose and tight registrations and analyzed their respective impacts on the colorization outcomes.

The outcomes are depicted in Figure 4.3. Figure 4.3 (a) and (b) illustrate that both pix2pix with tightly registered data and cycleGAN achieved comparable performance in the colorization process, displaying similar resolutions. This can be attributed to the relatively simple tissue structures, which are abundant in the training set and allowed both models to handle the colorization task effectively.

In Figure 4.3, a noteworthy difference is observed between the results produced by cycleGAN and pix2pix. In panels (c) to (g), we can observe hallucinations generated by cycleGAN, where random pink shades appear in regions where they do not correspond to any meaningful features in the ground truth images. These unrealistic colorizations are likely a consequence of the unsupervised nature of cycleGAN. In contrast, pix2pix exhibited more accurate colorization results, closely resembling the ground truth H&E images, thanks to its supervised learning approach with paired data. The distinct contrast emphasizes the advantage of using pix2pix for colorization tasks, as it can produce more faithful results by leveraging direct correspondences between input and output domains.

Our findings indicate that both loose and tight registration methods with pix2pix contributed to precise color estimation, but there was a noticeable variation in the resulting resolution and quality of colorization. Loose registration led to blurred results due to the inherent shift learned as a result of registration errors. On the other hand, tight registration provided more precise alignment, resulting in crisper images.

Additionally, in a previous study presented in [97], we demonstrated that in certain instances, utilizing loosely registered data with cycleGAN can lead to improved colorization outcomes. This finding is exemplified in Figure 4.4, where cycleGAN with loosely registered data (paired) produces colorizations that show enhanced performance compared to using unregistered (unpaired) data. This suggests that the impact of registration on colorization results can be context-dependent and may vary based on the specific characteristics of the dataset and the images being processed.

To evaluate the results, we utilized SSIM [126] and RMSE metrics to compare the

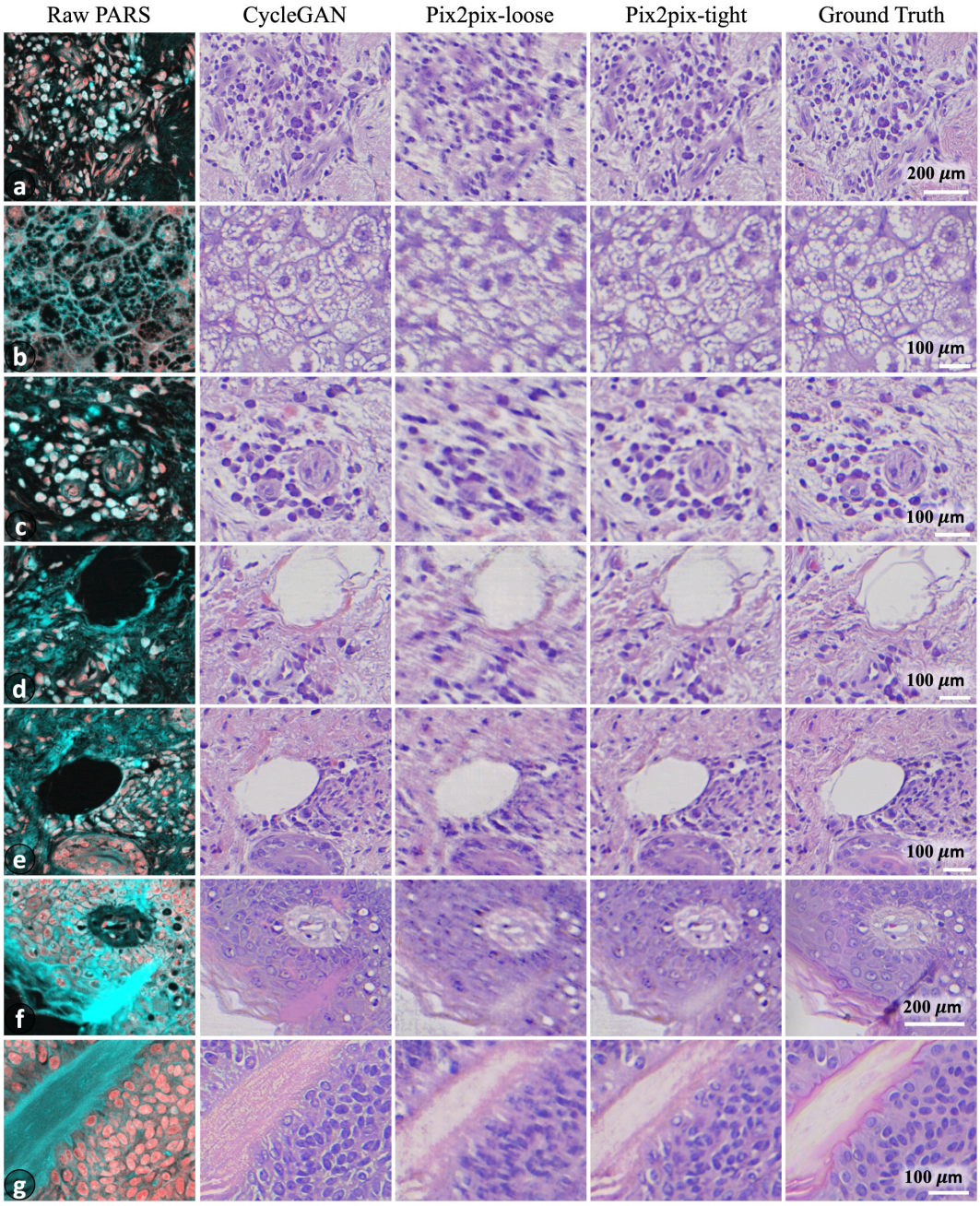



Figure 4.3: A comparison between the virtual staining results achieved by cycleGAN and pix2pix models using a human skin sample. (a)-(f) Different zoom-in field-of-views of the sample.

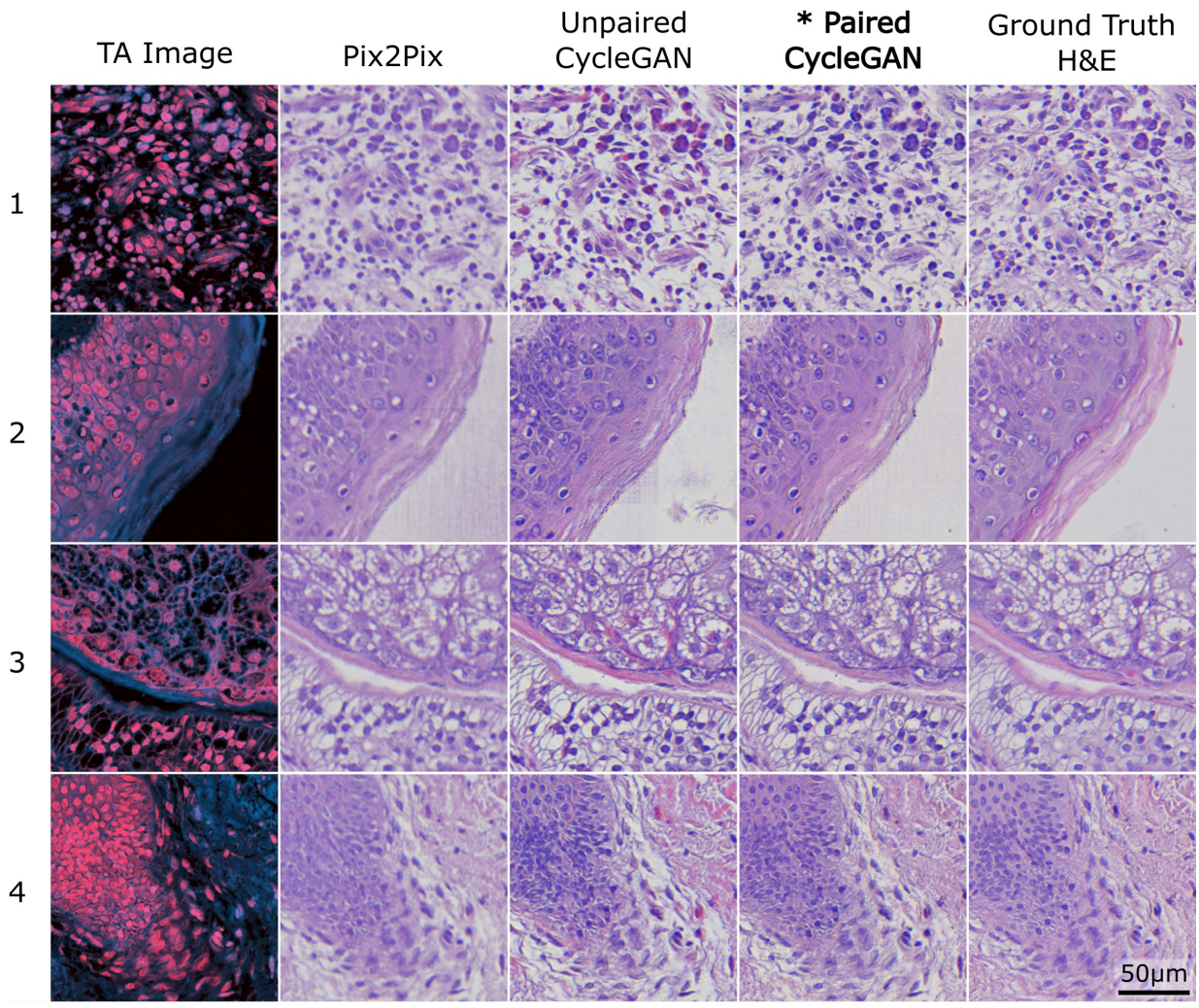



Figure 4.4: A comparison between the virtual staining results achieved by cycleGAN, with paired and unpaired data and pix2pix models using a human skin sample. Source: Reprinted from [97].

Table 4.1: Quantitative analysis summary of H&E colorization results on human skin dataset, using three models: cycleGAN, pix2pix with loosely registered data, and pix2pix with tightly registered data.

| Model | SSIM | RMSE |
|---|---|---|
| cycleGAN | 0.70 | 37.69 |
| pix2pix-loose | 0.56 | 47.96 |
| pix2pix-tight | 0.70 | 37.62 |

colorized images against the ground truth data for both pix2pix and cycleGAN.These metrics find extensive application, as demonstrated in Section 2.2, and are well-established, offering a consistent means of reporting and comparing outcomes. Furthermore, due to the availability of co-registered data, there was no necessity to employ metrics like FID and IS. The assessment outcomes are summarized in table 4.1. The statistical analysis indicates that cycleGAN and pix2pix with tightly registered data performed similarly, while, as anticipated, pix2pix with loosely registered data obtained the lowest score, showcasing the impact of registration on colorization accuracy.

The accuracy of pix2pix is attributed to its one-to-one correspondence, where it learns from paired data during training. This allows the model to grasp the pixel-level mapping between PARS and H&E images more effectively compared to cycleGAN, which employs an unsupervised learning approach and does not have direct correspondences between the input and output domains.

Both pix2pix and cycleGAN exhibit remarkable performance, but their requirements differ. Pix2pix necessitates near-perfect registration to achieve the best results, whereas cycleGAN shows promise with the potential to generalize well across large-scale datasets. The latter's ability to learn image-to-image translations without paired data makes it more adaptable to a wider range of applications, though it benefits from having access to a diverse and extensive dataset to achieve robust generalization.

## 4.4 Conclusion

In conclusion, this chapter delved into the application of PARS microscopy for virtual staining of unstained human skin tissue sections. We trained three distinct models: pix2pix with both loosely and tightly registered datasets, and cycleGAN with loosely registered data. Pix2pix has proven effective in producing digital representations of H&E stains, but its success relies on precise image alignment, which can be challenging due to registration inaccuracies and differences in imaging techniques between PARS and H&E. In contrast, cycleGAN addresses the limitations of paired data by handling unpaired image collections through cycle-consistency loss. While cycleGAN shows promise in certain scenarios, its lack of direct ground truth during training can lead to hallucinations and inaccuracies in virtual stains, affecting diagnostic reliability. Our research contributes valuable insights into the performance and requirements of both models, advancing virtual staining techniques and underscoring the need for ongoing refinement of GAN models to ensure their suitability in virtual staining and diagnosis applications.

# Chapter 5

# Multi-Channel Feature Extraction for PARS-to-H&E Translation

## 5.1 Motivation

The process of histological staining, particularly H&E staining, plays a vital role in the field of pathology for visualizing and distinguishing different tissue structures. Within the scope of virtual histological staining, PARS has demonstrated its effectiveness in generating H&E-like visualizations by directly capturing the H&E contrast [1,17,97], as discussed in Chapter 3. The non-radiative contrast observed in PARS resembles the contrast typically associated with hematoxylin stain, which predominantly targets cell nuclei. On the other hand, the PARS radiative channel primarily mimics the informational characteristics typically observed with eosin stain, which often highlights extracellular structures and connective tissue [17]. Despite this capability, the accuracy of colorization is not optimal. The model sometimes encounters difficulties in distinguishing structures of different tissue types, resulting in colorization errors. Therefore, further exploration is necessary to improve the virtual staining of PARS images. Here, we introduce a new concept aimed at improving the learning capacity of the colorization model by "separating" or "labeling" the biological structures before training the colorization model.

Upon closer examination of PARS data, it becomes apparent that the information contained within PARS, which encompasses structural details, is partially separated among its channels. As an illustration, the PARS non-radiative channels may reveal partial segmentation of nuclei, whereas the radiative channel primarily encompasses the tissue cytoplasm, red blood cells (RBCs), and connective tissue. Furthermore, previous studies [94,113] have

demonstrated that the shape of PARS non-radiative signals may be correlated with specific biological structures. Features from these signals can be extracted to facilitate the labeling and potentially enable the approximate separation of various biological structures.

Consequently, driven by the motivation that the structural information contained in PARS data inherently features the advantage of being partially separated among PARS channels, and it has been demonstrated [94, 113] that further separation is feasible, an experiment is developed to investigate the effect of data labeling or the separability of contrasts on colorization enhancement. The initial focus for concept validation is on utilizing H&E images exclusively. This approach allowed us to establish a solid foundation before extending our investigation to PARS data. By validating the concept using H&E images, we could gain insights into the effectiveness of contrast separability and its potential for enhancing colorization outcomes.

For the aforementioned initial validation, we focused on recoloring true H&E images by employing the separated H&E channels, shown in Figure 5.1 (a), as inputs to the virtual staining model. Our rationale behind the contrast separation approach is rooted in the belief that labeling the tissue structures by segregating them into multiple channels would aid the model in comprehending the underlying relationships between the input and target domains, ultimately leading to improved colorization performance. In order to establish a baseline for comparison, we also trained a model to recolor true H&E images using their grayscale versions (Figure 5.1 (a)).

To separate the H&E channels, we employed the sci-kit-image library [166] which follows the method proposed by Ruifrok et al. [167]. Subsequently, each separated channel is converted to grayscale and then merged to create an RGB image, with R representing the "H" stain, G representing the "E" stain, and the B channel duplicated from G. The dataset used in this experiment is extracted from a human skin sample, the same sample utilized in Chapter 6. The procedure for preparing the dataset is detailed in Section 6.2. To generate the training sets, overlapping patches of 256 $\times$ 256 pixels are extracted from the separated, grayscale, and color H&E images. About 1000 overlapping patches are split in a ratio of 80% for training and 20% for validation. The model used for training is the same one described in Section 3.5. The maximum number of epochs is set to 200 with an early stopping criterion to terminate the training when there is no further improvement in the generator loss. The trained model is then applied to the whole slide.

The results of this experiment are depicted in Figure 5.1 (b), clearly demonstrating that the utilization of separated channels yielded significantly improved colorization compared to using combined channels. When using the combined version as input, a noticeable blending of cell nuclei and connective tissue colorization occurred. However, with the

adoption of separated channels, this confusion is greatly minimized. This enhancement can be attributed to the distinct color characteristics associated with nuclei (shades of purple) and connective tissue (pink), and partially separating or labeling such structures facilitates more effective learning for the model and consequently results in more accurate colorization. Moreover, segregating information into multiple channels compared to having all information in a single channel, serves as a form of labeling and shows that having more information per pixel aids in better identification of structures and thus better estimation of colors. These compelling findings further solidify the driving factors behind the research presented in this chapter.

## 5.2 Introduction

Previously, the PARS microscope demonstrated its efficacy in generating virtual H&E stains using a pix2pix image-to-image translation model [1] as presented in Chapter 3. However, pix2pix demands accurate alignment of PARS and true H&E images at the pixel level. Perfect pixel-level registration is not always achievable due to alterations in the tissue during the staining process. Poorly aligned pairs weaken the colorization performance and lead to distorted or blurred translations [97], as illustrated in Chapter 4. To avoid such problems, we chose to employ a more flexible image-to-image translation model, cycleGAN [19] (introduced in Section 2.1.6) which is one of the most commonly used frameworks for unsupervised training [122]. While cycleGAN can be trained with unpaired data, in this chapter, registered pairs, which may not be perfectly aligned, are used as they help enhance the colorization results achieved by cycleGAN [97], primarily through mitigating hallucination artifacts.

Previous PARS virtual staining efforts used both the non-radiative and radiative channels from a single 266 nm UV excitation wavelength [1, 97, 164]. While these combined contrasts are highly analogous to the nuclear and connective tissue contrasts highlighted with H&E, some structures, like RBCs, are not strongly captured with UV excitation alone. An additional excitation wavelength can be used to target hemoglobin or RBCs to provide better separation between them and comparably sized features, such as nuclei. Providing additional information to the model may help it understand the separability between different structures and enhance its ability to learn the statistical relationship between the PARS and H&E domains.

Accordingly, an additional 532 nm excitation laser is employed in this work to help acquire RBC contrast, resulting in the non-radiative time-domain (TD) signals exhibiting two excitation peaks. The first peak, driven by a wavelength of 266 nm, is used

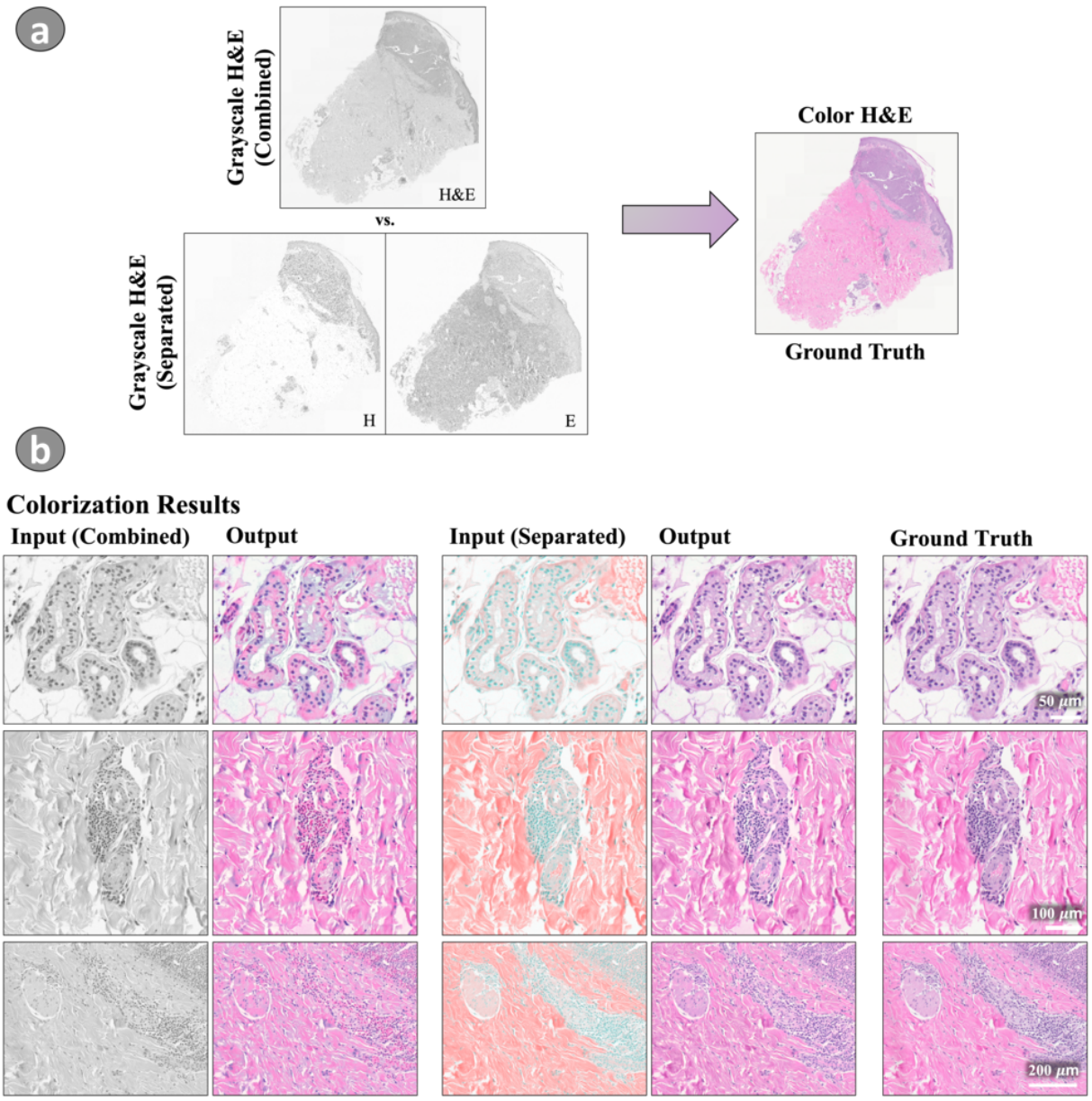



Figure 5.1: A visual representation demonstrating how the separation of H&E channels can contribute to improved colorization results. The experiment is done using human skin tissue. (a) Input channels: combined grayscale H&E versus separated grayscale H&E input, ground truth: color H&E. (b) Colorization results using the two input variations with their corresponding ground truth.

to specifically target DNA/RNA, while the second peak, from 532 nm, primarily targets hemoglobin (RBCs). Moreover, recently Pellegrino et al. [94, 113] showed that multiple features may be extracted from single peak time-resolved non-radiative relaxation signals. These features are extracted based on signal shape and may relate to more specific tissue structures [94, 113]. Hence, we choose to expand the number of input channels to the cycleGAN model by intelligently extracting features from the non-radiative TD signals to improve the distinction between different structures, following the motivation discussed in Section 5.1. To extract such features, we utilize a modified K-means (K*-means) method presented in [113]. These features are subsequently employed to reconstruct feature images, which, together with the conventional non-radiative and radiative images, form an array of image components. Such components can serve as inputs for the virtual staining cycleGAN model in different arrangements, as detailed in Section 5.9. The novelty of this work resides in the use of such features to train a cycleGAN model.

CycleGAN conventionally allows single- or three-channel input data [19, 168]. To effectively utilize the extracted features in virtual staining, this chapter introduces multi-channel cycleGAN (MC-GAN) that extends the existing cycleGAN model used in previous work [97] to support more than three channels. This approach allows for the extraction of multiple features, providing a better understanding of the potential of TD signals in improving virtual staining. The key contribution of this work is to enhance the performance of the colorization model by incorporating multiple features for labeling structures. Another contribution of this work is the introduction of a comprehensive pipeline that enables the extraction and selection of the most effective features to enhance the process of colorization.

## 5.3 Sample Preparation and Data Acquisition

The study utilized a dataset obtained from thin sections of formalin-fixed paraffin-embedded (FFPE) human skin tissues and mouse brain tissues. The human tissue samples are provided by the Alberta Precision Laboratories in Calgary, Alberta, Canada. The collection of these samples adhered to approved protocols established with the Research Ethics Board of Alberta (Protocol ID: HREBA.CC-18-0277) and the University of Waterloo Health Research Ethics Committee (Protocol ID: 40275). The requirement for patient consent was waived by the ethics committee, as samples are anonymized to remove all patient information, and tissues are archival samples not required for patient diagnostics. All experiments involving human tissues are conducted in compliance with the relevant guidelines and regulations of the government of Canada, such as "Ethical Conduct for Research Involving

Humans (TCPS 2)". The mouse brain samples were prepared at the National Institute of Health in Bethesda, Maryland, United States. These samples adhered to approved protocols under the University of Waterloo Health Research Ethics Committee (Protocol ID: 44595).

The preparation of unstained tissue sections involves several steps. Initially, tissue is resected and immediately fixed in a 10% neutral buffered fixative for up to 48 hours. Next, the sample is submerged in a series of alcohol exchanges to dehydrate the tissue. Dehydration is followed by a series of rinses in xylene, a tissue-clearing agent which removes any fat residues. The dehydrated and cleared tissue is then soaked in molten paraffin to allow for the complete infiltration of paraffin into the tissue. Finally, tissues are embedded in paraffin and allowed to solidify at room temperature, forming an FFPE tissue block. Tissue sections are prepared by using a microtome to slice thin sections of about 3-5 µm from the FFPE block. Thin sections are then transferred to a glass microscope slide by means of a water bath and allowed to dry. Slides are heated to 60 °C for 60 minutes in a laboratory to evaporate excess paraffin. After this process, the thin sections are ready for imaging using the PARS system, as outlined in Section 5.4. Once the PARS dataset is acquired, the thin tissue slides are stained using H&E. The stained slides are completed with a mounting medium and a coverslip. The stained sections are imaged at 40x (Morpholens 1, Morphle Digital Pathology), thereby generating the corresponding H&E ground truth dataset. The virtual staining procedure follows the method described in Section 3.1.

## 5.4 PARS Imaging

While this work does not focus on the PARS system design, introducing the system is important for understanding the data collection process. For a more detailed exploration of the system architecture and image formation process, refer to the work done by Tweel et al. [164]. Briefly, the experimental setup is depicted in Figure 5.2. This architecture features two excitation sources: a 266 nm UV laser (Wedge XF, RPMC) and a 532 nm visible laser (Wedge XF, RPMC). The detection source is a 405 nm OBIS-LS laser (OBIS LS 405, Coherent). The excitation and detection sources are combined using a dichroic mirror and focused onto the sample using a 0.42 NA UV objective lens (NPAL-50-UV-YSTF, OptoSigma). This configuration provides a maximum spatial resolution of approximately 400 nm. The detection light, containing the optical scattering and non-radiative relaxation is collected after transmission through the sample using a second objective lens (100X Mitutoyo Plan Apo, Mitutoyo). The same lens also collects the radiative emissions from the sample. To separate the non-radiative detection light from the radiative relaxation, a

spectral filter (NF405-13, Thorlabs) is employed. Then, the radiative emission amplitudes are directly measured using a photodiode (APD130A2, Thorlabs). Concurrently, the non-radiative detection light is directed to another photodiode (APD130A2, Thorlabs) where the optical scattering and the absorption-induced intensity modulations are captured.

PARS images are generated through the following process. The sample is scanned over the stationary objective lens using mechanical stages. The 266 nm and 532 nm excitation sources are pulsed continuously at a rate of 50 kHz, with energies of 150-200 pJ, and 2.5 nJ for the 266 nm and 532 nm sources, respectively. Concurrently, the continuous wave detection laser operates at 3-5 $\mu$ W during imaging. The 532 nm pulses are synchronized to occur 500 ns after the 266 nm pulses. The stage motion is then tuned to ensure the 266 nm excitation event occurs every 250 nm. During each excitation event, the optical scattering, and radiative and non-radiative relaxation signals are collected from the corresponding photodiodes, for each excitation wavelength. Additionally, a position signal is recorded from the scanning stages. The radiative signals are condensed into a single feature by extracting the amplitude of the measured signal. The optical scattering is extracted by averaging the transmitted detection prior to excitation. The post-excitation, time-resolved non-radiative relaxation modulations are stored in their entirety for further processing.

To generate a baseline model for comparisons, the non-radiative signal reconstruction method used in previous embodiments is employed [1]. This method involves signal extraction by integrating the post-excitation modulation energy of the non-radiative signals. The baseline is used to contrast the efficacy of intelligently extracting features (explained in Section 5.6.1) from the non-radiative signals for enhancing the virtual staining process.

## 5.5 Image Registration and Preprocessing

In this chapter, we chose to use registered PARS and H&E pairs for training a cycleGAN for virtual staining, as discussed in Section 5.2. However, the PARS and H&E images are not inherently co-registered due to the H&E staining procedure and the two different acquisition processes involved. Subsequently, additional steps, including field-of-view matching and registration, are required to align the PARS and H&E images for model training and performance analysis.

The preprocessing and registration process follows the workflow previously described by Boktor et al. [1]. This includes extracting small fields from the PARS and H&E images and then coarsely matching them for one-to-one registration. The control point registration tool from the MATLAB Image Processing Toolbox is used for registration, with PARS as

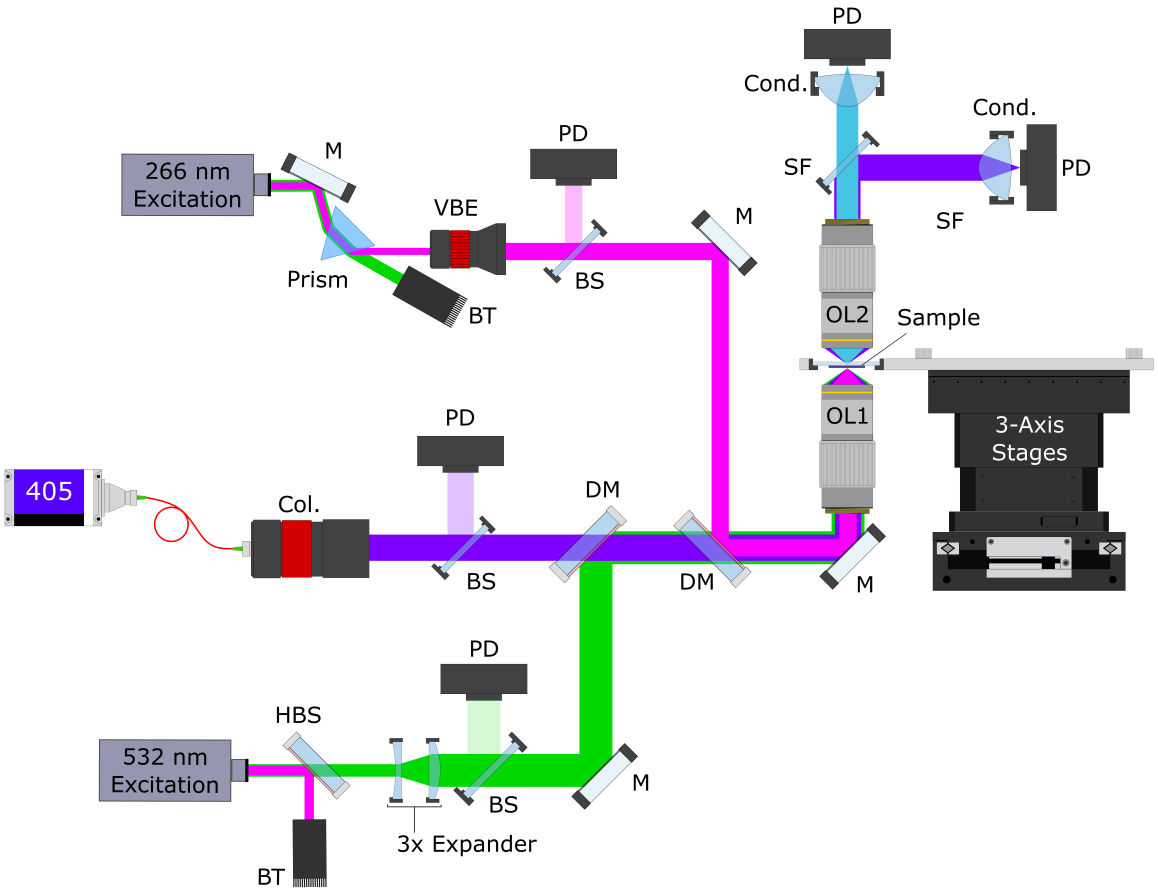



Figure 5.2: Simplified PARS histology optical architecture. Component labels are defined as follows: mirror, M; dichroic mirror, DM; variable beam expander, VBE; collimator, Col.; condenser lens, Cond.; spectral filter, SF; beam sampler, BS; photodiode, PD; objective lens, OL; harmonic beam splitter, HBS; beam trap, BT. Source: Reprinted from [2].

the reference images and H&E as the registered images. Control points are manually selected and refined to minimize distortions. A non-rigid geometric transformation [169] is then fitted between the two images, and the transformation is applied to the H&E images, resulting in co-registered pairs of PARS and H&E images. Following registration, PARS images are preprocessed as follows. Each image is normalized where each pixel is divided by its maximum possible value, then the contrast is enhanced by saturating the highest and lowest 1% of pixel values. Finally, a color reversal is performed by subtracting each pixel from the maximum pixel value to match the PARS images with the grayscale ground truth images. The same preprocessing is applied to all datasets used in this chapter.

## 5.6 Time-Domain Feature Extraction

### 5.6.1 K*-means Clustering

To extract material-specific information from non-radiative TD signals, a method is needed to identify constituent time-domain features that accurately represent the underlying tissue target. The method, proposed by Pellegrino et al. [113], is utilized here which features K-means clustering with a modified approach to compute cluster centroids. To generate feature images from the non-radiative TD signals, two steps are involved: feature learning and feature extraction.

The intelligent clustering aims to identify $K$ characteristic shapes in the signals, described as a set of $K$ centroids, $\mathcal{F} = \{f_i(t)\}, i = 1, \ldots, K$. TD signals are treated as Cartesian vectors in space $\mathbb{R}^n$, where $n$ corresponds to the number of samples in each TD signal, and thus the shape of the signal is associated with the angle of the corresponding vector and the distance between TD signals is quantified by the sine of the angle between them, resulting in a maximum distance for orthogonal signals and zero distance for scaled or inverted signals. Cluster centroids are computed as the principal component of the combined set of each cluster and its negative, ensuring that the learned centroids are resilient to noise.

Following the K*-means clustering, a set of feature vectors $\mathcal{F} = \{\vec{f_i}\}$ is obtained, which can represent the signals as a weighted sum. These feature vectors are then arranged in the form of a matrix of features, $F = [\vec{f_1}|\vec{f_2}|\ldots|\vec{f_K}]$. The amplitudes of the learned features are extracted by transforming TD signals from the time domain to the feature domain. This is performed by multiplying each TD signal with the pseudo-inverse of $F$ [113, 170]. The result is an array of $K$ feature images, $M_f = [m_{f_1}, m_{f_2}, \ldots, m_{f_K}]$. Figure 5.3 shows an

example of the feature images at $K = 3$ using the human skin and mouse brain datasets, combined to produce a color image for better visualization. It is important to note that K*-means clustering is applied separately to each dataset. The resulting color images demonstrate distinct coloring for different tissue structures, such as RBCs (appear in red in human skin images), cell nuclei (mostly green), and various types of connective tissue (shades of dark purple and red). This observation highlights the richness of information contained within the non-radiative TD signals, extending far beyond a single brightness value. The incorporation of extracted features plays a crucial role in effectively labeling tissue structures that exhibit different signal shapes.

The appropriate value of $K$ is determined according to the specific dataset as discussed in Section 5.9. A minimum of 2 clusters is examined during the feature extraction process. However, it is important to impose an upper limit on $K$, set to 6 in this work, to prevent the generation of redundant clusters and avoid the introduction of visually indistinguishable or uninformative features. Following the determination of the optimal value for $K$, a set of $K$ feature images is generated. These feature images alongside the conventional non-radiative and radiative image components form an array of images, which can serve as the inputs for the colorization model in different combinations.

### 5.6.2 Principal Component Analysis (PCA)

As discussed in Section 2.1.7, PCA is designed to uncover the primary components that account for the highest variance within the data. However, it does not directly derive meaningful biological features from the time-domain signals [113]. This limitation becomes particularly noticeable in imaging techniques like PAM, where the signals often exhibit complex shapes and patterns. Although PCA is proficient at reducing data dimensionality by prioritizing variance, it may fall short in adequately depicting the intricate characteristics and distinguishing various structures within highly complex signals.

Nonetheless, the importance of being able to distinguish between different structures is mainly relevant in specific contexts, such as virtual labeling. In the context of our virtual staining application, the situation may differ. Given our objective of segregating data into multiple channels and diminishing redundancy, individual Principal Components (PCs) may not inherently embody biologically meaningful features, however, they could serve as valuable inputs for training virtual staining models. Hence, we opted to explore the utilization of PCA features for our virtual staining task.

We applied a similar procedure to PCA as we did with K*-means, with the distinction being that we substituted centroids with PCs. To ensure a fair comparison with K*-means,

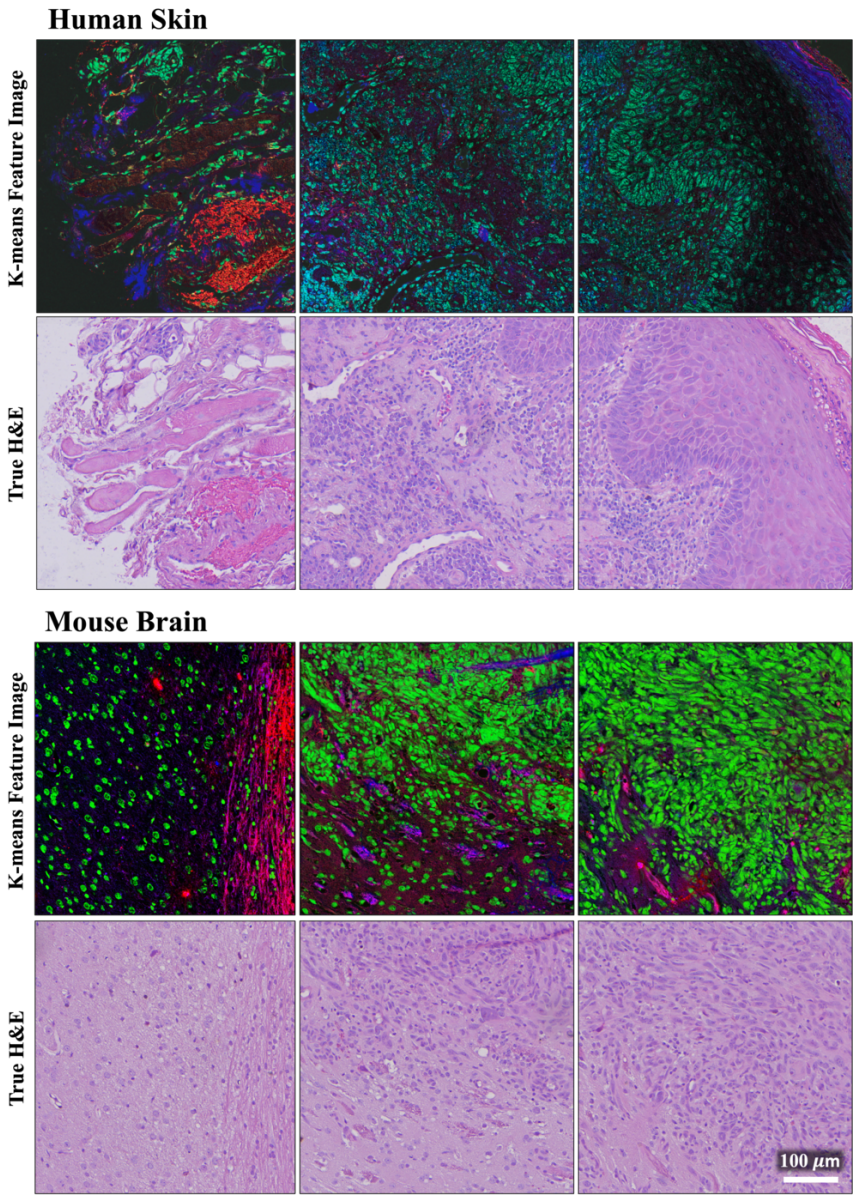



Figure 5.3: The figure illustrates examples of feature images extracted from the NR TD signals at $K = 3$ using the K*-means method, along with their corresponding true H&E images. The feature images are combined to generate color images. These images reveal remarkable tissue differentiation.

our model training encompassed an exploration of the optimal number of PCs, ranging from 2 to 6, to achieve the highest performance.

## 5.7 Multi-Channel GAN (MC-GAN)

In this chapter, cycleGAN is applied to convert label-free PARS images into virtually stained images that resemble their corresponding H&E histochemically stained samples. CycleGAN learns feature relations between an input image domain, $A$, and a target image domain, $B$, and generates the generators $G : A \rightarrow B$ and $F : B \rightarrow A$ and the discriminators $D_A$ and $D_B$. The generators aim to generate realistic images that resemble the target domain while preserving the essential characteristics of the input domain. The discriminator, on the other hand, tries to differentiate between real images from the target domain and fake images produced by the generator. The discriminator provides feedback to the generator by assessing the fidelity of the generated images, enabling the generator to improve its output quality over time through an adversarial training process.

In this work, domain $A$ is PARS imagery while domain $B$ is H&E imagery. CycleGAN is able to operate in a scenario where paired examples are not available, nevertheless, the absence of paired samples poses challenges in mapping between the source and target domains as it is under-constrained. To tackle this, cycleGAN incorporates an inverse mapping and introduces a cycle consistency loss [19]. This loss enforces that the translated images can be reliably reversed to their original form. However, it is worth noting that although paired examples are not required for training, the utilization of paired samples in our cycleGAN training does lead to visible performance improvements [97].

Typically, cycleGAN models are trained using single- (grayscale) or three-channel (RGB) images [14,19,105]. In previous works [1,97,164], these RGB channels are directly replaced with the non-radiative and radiative channels, which posed no issues as the number of channels did not exceed three. However, the focus of this research is to improve the efficacy of the virtual staining process using intelligently extracted features from non-radiative PARS signals, resulting in a scenario where the colorization model has more than three input channels. It is important to note that the use of three channels in conventional cycleGAN models is solely due to the nature of working with RGB images, and there is no inherent reason to restrict the input channels to three in data-fusion methods. In fact, the optimal number of input channels is likely to differ from three. Hence, the proposed MC-GAN model enables the incorporation of additional features by expanding the number of input channels beyond three. The integration of extra channels in the MC-GAN model allows for a wider range of information to be utilized during training. This extension enhances

the model's ability to capture and leverage diverse information, which can potentially lead to improved colorization performance.

In normal cycleGAN, $N$-channel input generates $N$-channel output. However, our target H&E domain is an RGB image with three channels. Only a three-channel output is required, regardless of the number of channels in the source domain. To allow a three-channel output with an $N$-channel input, we duplicate the last channel ($B$) in the target domain to expand the target H&E domain dimensions to match the source PARS domain. When extracting the colorization results, we discard these duplicated channels. Except for the modifications of the multi-channel input and output, the architecture utilized in this chapter remains identical to the original cycleGAN [19].

## 5.8 Training Settings

Two datasets are employed in this work: human skin and mouse brain data. To generate the training sets for the two datasets, overlapping patches of 256 $\times$ 256 pixels are extracted from the PARS and H&E images. For the human skin dataset, approximately 500 overlapping patches are extracted from one tissue slide sample, while for the mouse brain dataset, benefiting from a larger available sample, around 2000 overlapping patches are extracted. The pairs are split in a ratio of 70% for training, 10% for validation, and 20% for testing. Following the same values from Section 4.2.4, the learning rate is set to 0.0002, and the maximum number of epochs is set to 200 with an early stopping criterion to terminate the training when the generator loss stops improving. The trained model is then applied to the test images from unseen data extracted from a different sample. The test images are also subdivided into overlapping patches of 256 $\times$ 256 pixels. Empirically, we found that an overlap of $\sim$ 50% is usually sufficient to avoid visible artifacts at the borders of adjacent patches in the final stitched image. The colorization algorithm is implemented in Python version 3.10.6 and model training is implemented using PyTorch version 2.0.0 with the support of CUDA version 12.

## 5.9 Results and Discussion

In previous PARS virtual staining embodiments [1, 97, 164], only the non-radiative and radiative channels (denoted as NR and R, respectively, in this section for notation simplicity) are used as inputs for the virtual staining model. Traditionally, NR images are created by integrating the post-excitation modulations into the detection signal. This method omits

valuable temporal information because the shape of the signal may contain information associated with specific biological structures [94, 113]. In this chapter, we opt to broaden the range of input channels by incorporating features extracted from the NR TD signals. These features augment information at each pixel location, which can enhance the colorization model's ability to comprehend the statistical transformation between the input and target domains.

The multi-channel virtual staining workflow is shown in Figure 5.4. The pipeline consists of two main parts: (1) feature learning, extraction, and selection, and (2) MC-GAN model training. Feature learning takes place using a representative subset of the NR TD signals. Feature extraction is performed on all the TD signals to form feature images. These different feature images, along with the NR signal integral (as described in Section 5.4) from each excitation wavelength (266 nm and 532 nm), and the R channel from 266 nm excitation only, are then fed into the feature selection stage. Feature selection is used to enhance the model's prediction power by eliminating redundant data, increasing the contrast between the selected features, and reducing the training volumes and times. Finally, the images of the selected features are used as input to the proposed MC-GAN model, and the true H&E serves as the ground truth.

### 5.9.1 Feature Extraction

#### K*-means

A study (the "K-study") is conducted to determine the optimal number of features $K$ to extract from each section. In this K-study, feature extraction is performed using the K*-means algorithm (presented in Section 5.6.1) generating $M_f^K$ for every $K \in \{2, \ldots, 6\}$. The feature extraction results for each $K$ are shown in Figure 5.5. It can be observed that when $K < 3$, the available features are insufficient to effectively distinguish and separate significant structures into distinct feature images. For example, when $K = 2$, structures such as connective tissue and cell nuclei are merged into a single feature image. However, when the value of $K$ is increased to 3, these structures are successfully separated into distinct feature images. It can also be noticed that once the number of features reaches more than three, the additional feature images start to lose their significance because features or patterns of the same kind are compelled to separate into two different feature images. This phenomenon can be observed with connective tissue (highlighted with yellow arrows) when $K = 4$ as well as with RBCs (highlighted in red) and nuclei (highlighted in green) when $K = 5$. These observed patterns are further supported by conducting

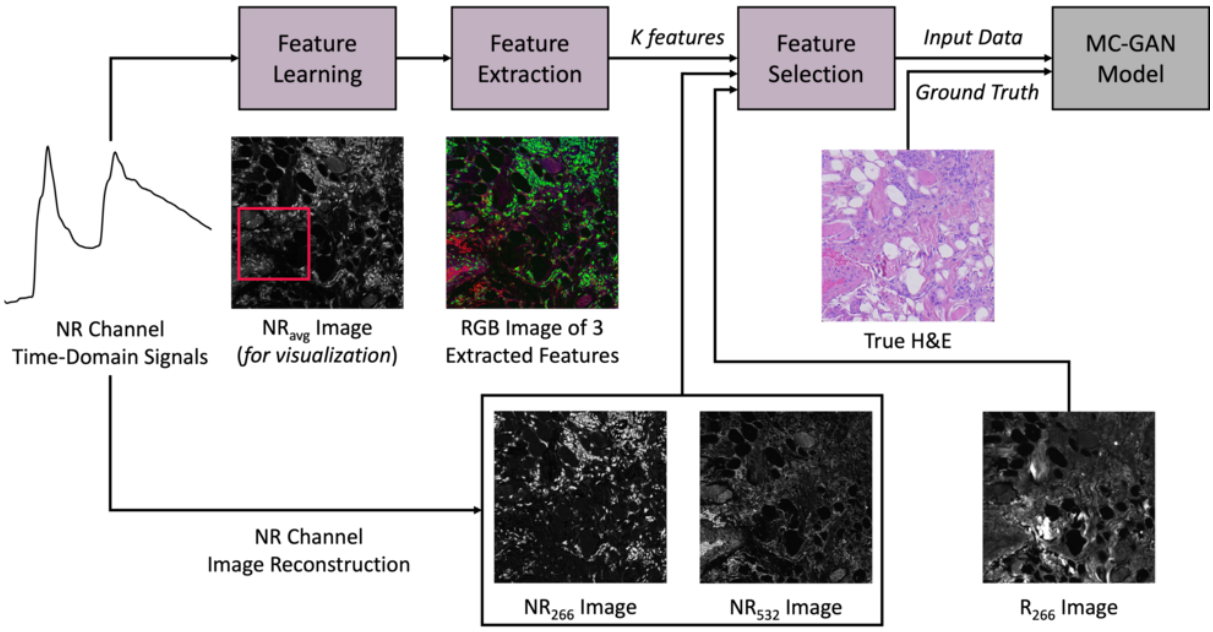



Figure 5.4: Multi-channel virtual staining workflow. First, feature learning, of $K$ features, takes place using a subset (shown in red box) of the NR channel TD signals. Second, feature extraction is performed on all the TD signals of the data in hand forming $K$ feature images. NR images of each excitation wavelength (266 nm and 532 nm in this case) and R images are extracted separately and passed along with the $K$ feature images to the feature selection phase. The selected features are then used as the input data to the proposed MC-GAN model, and the true H&E is used as the model ground truth.

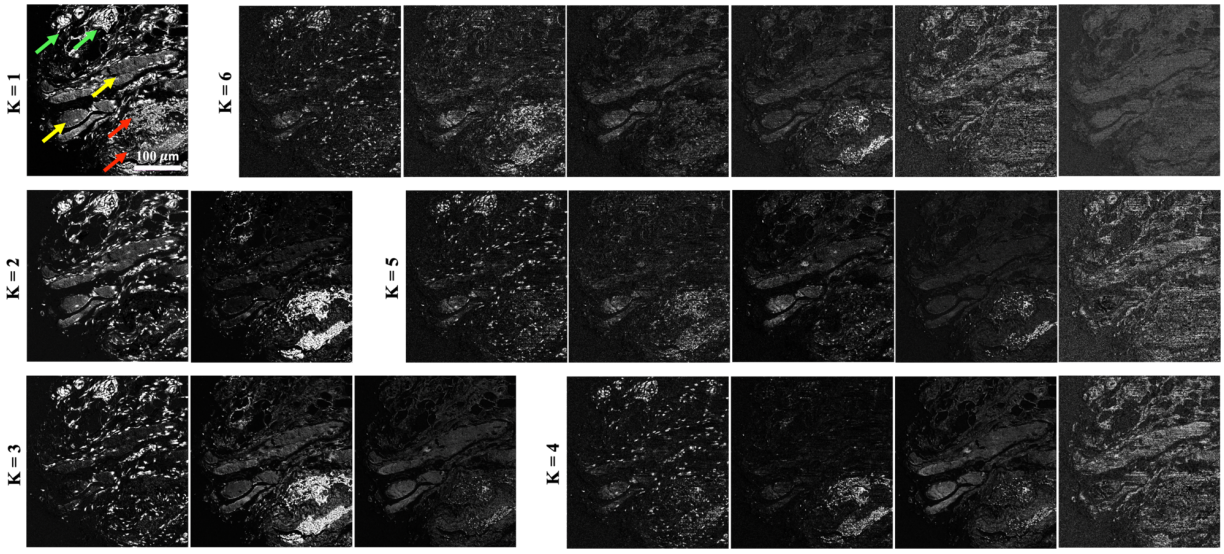



Figure 5.5: Feature extraction results of the K-study. When $K < 3$, major structures like RBCs (red arrows) and connective tissue (yellow arrows) are not effectively separated. When $K > 3$, the features start to become redundant and are forcefully separated like in the case of RBCs and cell nuclei (green arrows).

experiments where these features are utilized as input for the virtual staining process. The details and outcomes of these experiments will be elaborated upon later in this section.

Afterward, the MC-GAN is trained using the R channel and $M_f^K$ for each $K$. Since the $K$ features are extracted from the NR channel TD signals, they are independent of the R contrast. Hence, the R channel is always used as an additional component in the virtual staining phase of the K-study for a fair comparison. The model performance is then assessed.

Visual assessment, and Structural Similarity Index (SSIM) [126,171], computed between the colorized images and their corresponding ground truth, are used to determine the optimal value of $K$. The best $K$ is selected independently for the two datasets in hand. For the human skin dataset, the K-study revealed that the best results are obtained when $K = 3$, as shown in Figure 5.6, whereas, for the mouse brain dataset, the K-study SSIM results indicate an optimal value for $K$ of 2.

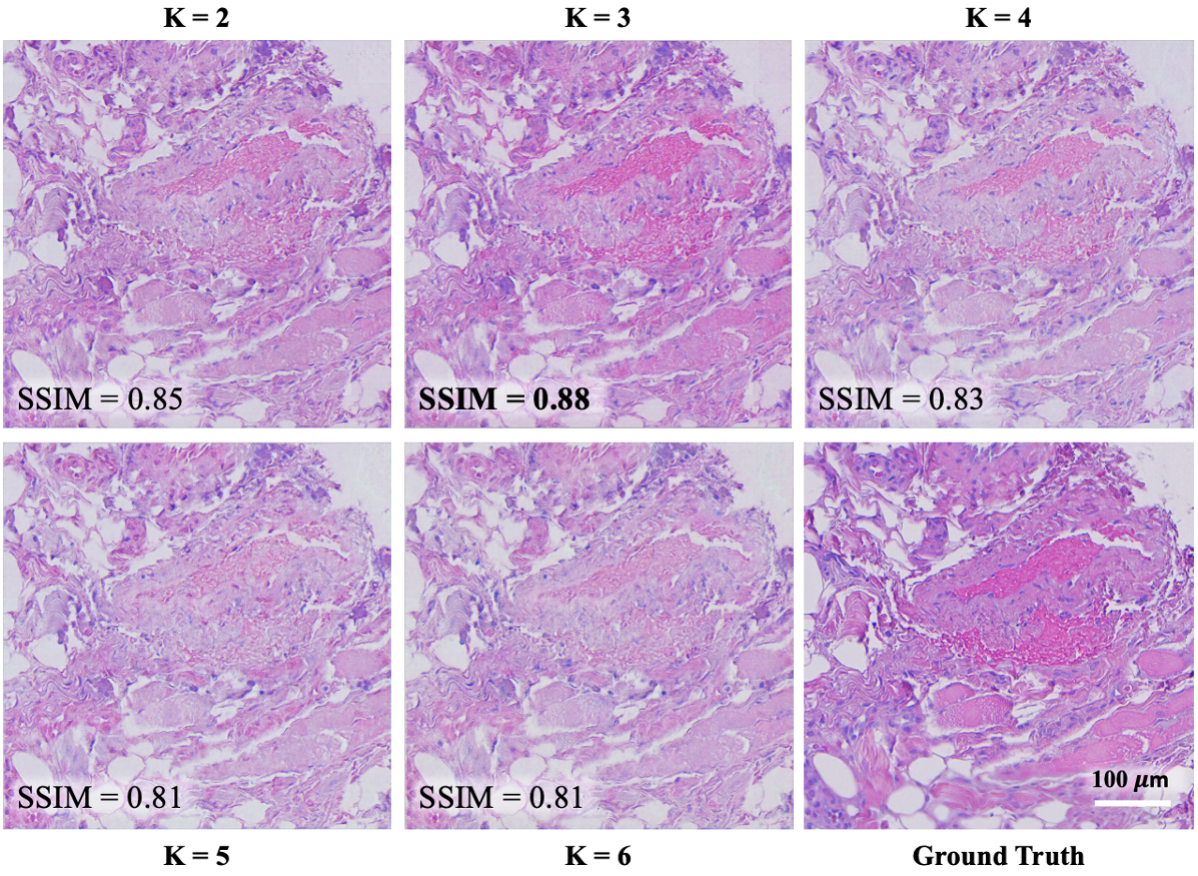



Figure 5.6: Virtual staining results of the K-study for each $K \in \{2, \ldots, 6\}$ using human skin data. The findings indicate that the optimal colorization is produced when $K = 3$.

### PCA

Similar to the K-study, PCA is conducted on the PARS signals, followed by an investigation to identify the most suitable number of PCs. As mentioned earlier in this section, the first six PCs are selected resulting in the creation of six distinct feature images, each corresponding to an individual PC, as depicted in Figure 5.7. Following that, we trained five MC-GAN models, incorporating the first $'X'$ components for each $X$ in the range of 2 to 6, in addition to the R channel, mirroring the approach employed for K*-means. The outcomes of each model can be observed in Figure 5.8.

Once more, we employed a visual evaluation and computed the SSIM between the colorized images and their respective ground truth to ascertain the ideal number of PCs. Notably, PCA is initially performed using solely the same human skin dataset for the purpose of comparison with the K*-means approach.

The results indicate that the optimal number of PCs is three. However, it is important to note that the first three PCs account for only 60% of the variance in the data, whereas the initial 36 PCs explain 70%, and the first 137 PCs explain 90%. Therefore, using a limited number of PCs to represent the information contained within the PARS signals might not adequately encompass the crucial characteristics of the data, making it a less practical approach.

Moreover, the results demonstrate that in terms of segmenting biological features, particularly for virtual labeling purposes, K*-means outperforms PCA. The PCA features did not reveal meaningful attributes that could be associated with specific biological structures, as illustrated in Figure 5.7. To illustrate, $PC_1$ encompasses all distinct biological structures with varying intensities, whereas $PC_2$ duplicates one of these structures, specifically cell nuclei. Furthermore, $PC_5$ and $PC_6$ exhibit nearly identical biological structures. Consequently, K*-means proves to be a more suitable choice for the task of virtual labeling.

However, concerning virtual staining, PCA exhibited commendable performance, as shown in Figure 5.8, with similarity metrics results comparable to those of K*-means. Nevertheless, visually, K*-means notably outperformed PCA, particularly in aspects like contrast, exemplified by the enhanced contrast of features such as RBCs. These outcomes suggest that K*-means may be a more suitable choice for processing PARS data, not only for virtual labeling, as elaborated in [113], but also for virtual staining applications. Given these considerations, we have decided to advance with K*-means features for the subsequent phase.

Finally, we conducted an ablation study to assess the impact of integrating both K*-means and PCA features with the NR and R conventional channels. Specifically, we

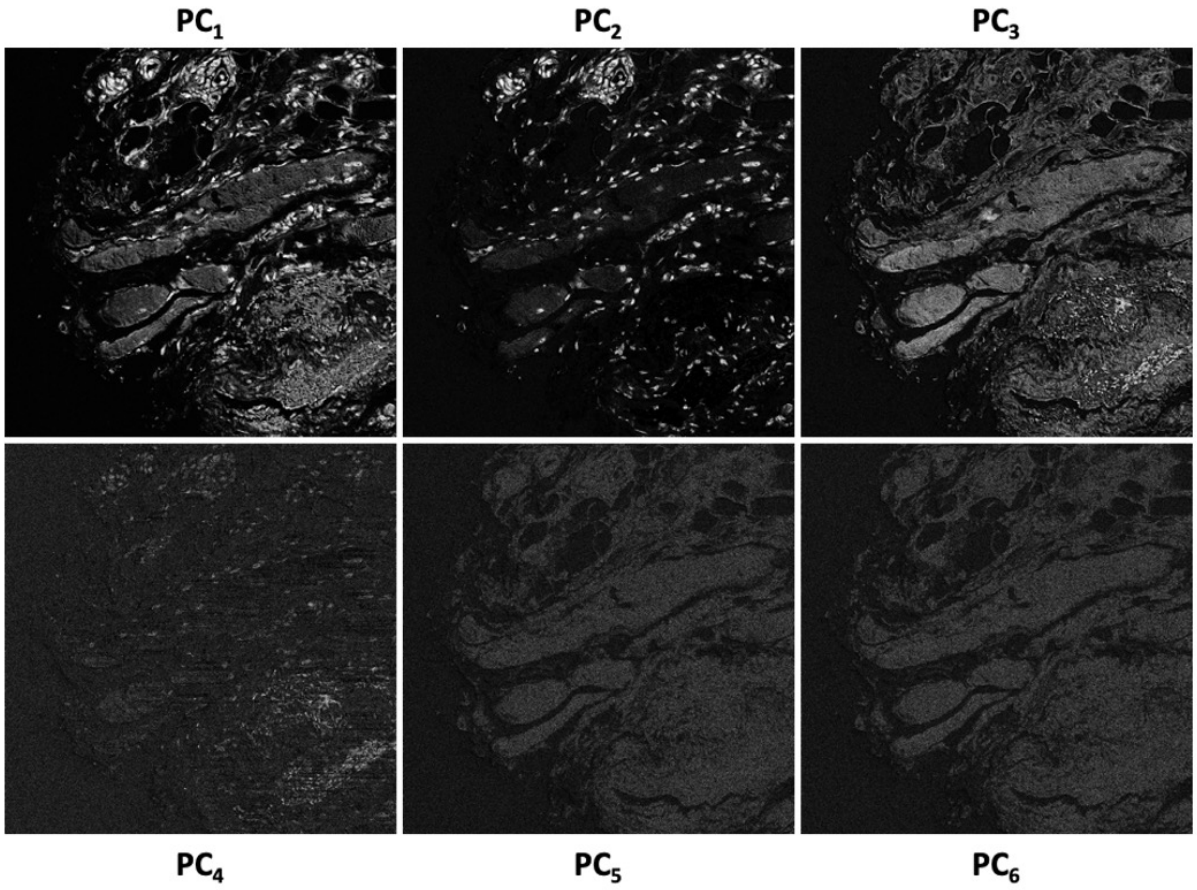



Figure 5.7: Feature extraction results using PCA for the first $'X'$ components for each $X \in \{2, \ldots, 6\}$ using human skin data. It can be noted that the PCA features do not align with identifiable biological structures.

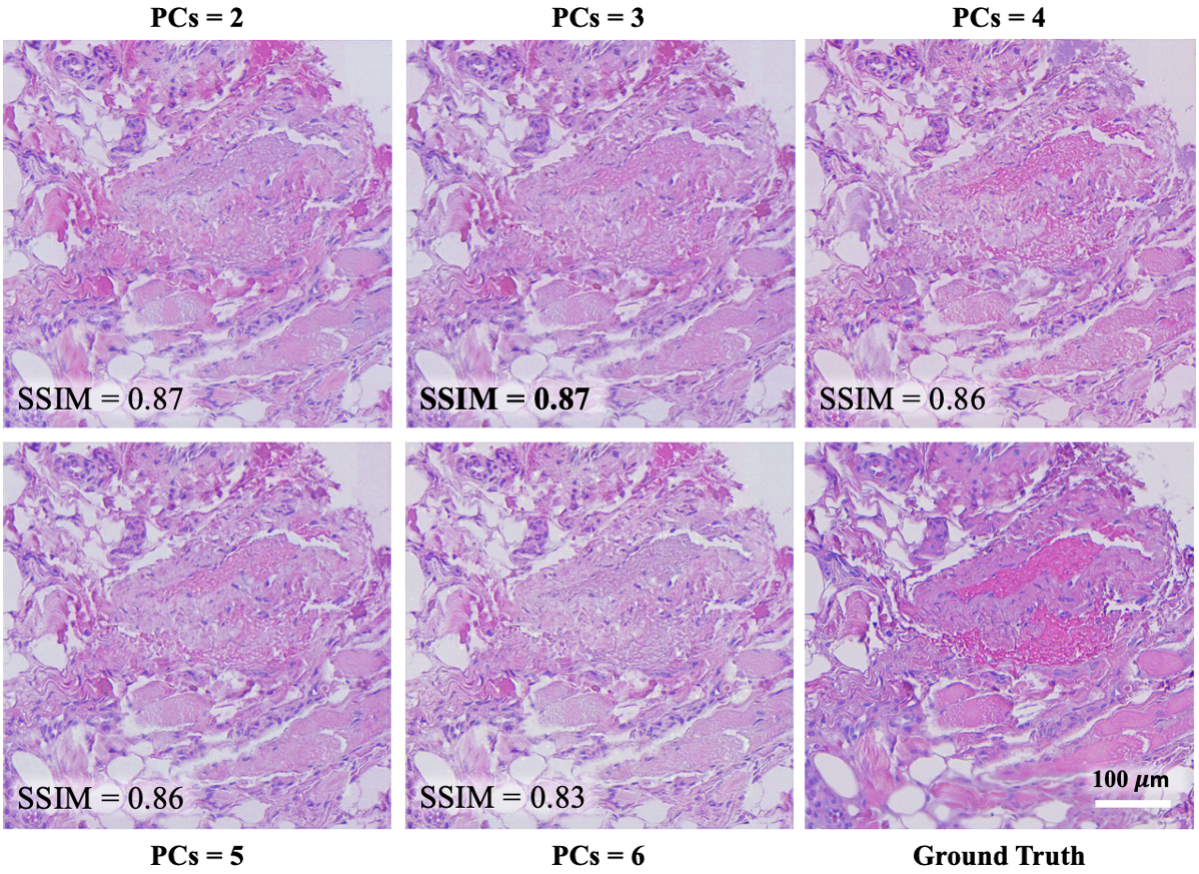



Figure 5.8: Virtual staining results using PCA for the first $'X'$ components for each $X \in \{2, \ldots, 6\}$ using human skin data. The findings indicate that the optimal colorization is produced when $X = 3$.

Table 5.1: Ablation study for feature extraction using K*-means and PCA features, showing that K*-means features are superior to PCA and conventional features for the purpose of virtual staining.

| **Input Features** | **Description** | **SSIM** |
|---|---|---|
| $NR_{532}$, $NR_{266}$, $R_{266}$, $m_{f_1}$, $m_{f_2}$, $m_{f_3}$ | Conventional + 3 K*-means Features | 0.87 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $m_{f_2}$, $m_{f_3}$ | (-$m_{f_1}$) | 0.87 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $m_{f_1}$, $m_{f_3}$ | (-$m_{f_2}$) | 0.86 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $m_{f_1}$, $m_{f_2}$ | (-$m_{f_3}$) | **0.88** |
| $NR_{266}$, $R_{266}$, $m_{f_1}$, $m_{f_2}$, $m_{f_3}$ | (-$NR_{532}$) | 0.86 |
| $NR_{532}$, $R_{266}$, $m_{f_1}$, $m_{f_2}$, $m_{f_3}$ | (-$NR_{266}$) | 0.87 |
| $NR_{532}$, $NR_{266}$, $m_{f_1}$, $m_{f_2}$, $m_{f_3}$ | (-$R_{266}$) | 0.85 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $PC_1$, $PC_2$, $PC_3$ | Conventional + 3 PCA Features | 0.86 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $PC_2$, $PC_3$ | (-$PC_1$) | 0.85 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $PC_1$, $PC_3$ | (-$PC_2$) | 0.85 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $PC_1$, $PC_2$ | (-$PC_3$) | 0.86 |
| $NR_{266}$, $R_{266}$, $R_{266}$, $PC_1$, $PC_2$, $PC_3$ | (-$NR_{532}$) | 0.86 |
| $NR_{532}$, $R_{266}$, $R_{266}$, $PC_1$, $PC_2$, $PC_3$ | (-$NR_{266}$) | 0.83 |
| $NR_{532}$, $NR_{266}$, $R_{266}$, $PC_1$, $PC_2$, $PC_3$ | (-$R_{266}$) | 0.83 |
| $NR_{532}$, $NR_{266}$, $R_{266}$ | Conventional Features Only | 0.87 |
| $NR_{532}$, $NR_{266}$ | (-$R_{266}$) | 0.81 |
| $NR_{532}$, $R_{266}$ | (-$NR_{266}$) | 0.87 |
| $NR_{266}$, $R_{266}$ | (-$NR_{532}$) | 0.85 |

utilized features derived from $K = 3$ and $PCs = 3$, which were found to produce the best results in the earlier study. We trained individual models using conventional channels alone, conventional channels combined with K*-means features, and conventional channels combined with PCA features. We systematically removed features one by one to assess their influence on the quality of colorization. The summarized results can be found in Table 5.1.

In general, the combination of K*-means with the conventional channels resulted in the highest SSIM scores, while incorporating PCA features led to the least favorable outcomes, even worse than using conventional channels alone. Additionally, the results using K*-means indicated that removing $m_{f_3}$ yielded improved results, suggesting that certain combinations of features could potentially confuse the model and negatively affect its performance. Regarding PCA, employing all three PCs alongside the conventional channels, in contrast to excluding the PCs, yielded the most favorable outcomes. Nevertheless, with

this combined approach, the results remained inferior to those achieved when solely utilizing the conventional channels. Overall, these results reinforced the decision to employ K*-means features in the following phase, which involved the C-study, for selecting the optimal feature combination.

### 5.9.2 Feature Selection

After extracting the features, a combinatorial study (the "C-study") is conducted to identify the optimal subset of K*-means features. In the C-study, the analysis encompasses all the available image components. By utilizing the optimal value of $K$ obtained from the K-study, a set of features denoted as $M_f^{opt}$ is generated. Additionally, the conventional NR and R channels are combined with $M_f^{opt}$ feature images forming an array of images $A$, where $A = [NR_{532}, NR_{266}, R_{266}, M_f^{opt}]$. This array provides a comprehensive representation of the image data. It is possible that using all the elements in $A$ to train a model could lead to redundancy and thus confuse the model. Consequently, the C-study conducts an exhaustive search across all the possible combinations of elements in array $A$ to determine the optimum feature combination for creating a robust colorization model. The size of array $A$ is determined by the sum of the number of features obtained from the K-study and the three conventional channels. Therefore, the total number of elements in array $A$ is given by $N = K + 3$. The number of possible combinations is $2^N - 1$.

To evaluate the model performance for the C-study, pixel-wise evaluation metrics, SSIM, Peak Signal-to-Noise Ratio (PSNR) [127], and Root Mean Squared Error (RMSE), are calculated since paired input and ground truth images are available [14]. The three metrics are computed between the colorized images and the true H&E. Images of the two domains are blurred prior to computing the metrics to avoid the effect of registration errors. Based on the assessment outcomes, a feature selection process is carried out to determine the best-performing members from set $A$, as presented in Section 5.6.1. These are then used to train virtual staining models. The results of the two datasets are presented in the following subsections.

### 5.9.3 Virtual Staining Results

#### Human Skin Dataset

The number of elements in $A$ is determined based on the value of the best $K$ that is selected during the K-study. For instance, using $K = 3$ results in an image set $A =$

$[NR_{532}, NR_{266}, R_{266}, m_{f_1}, m_{f_2}, m_{f_3}]$ which contains three extracted feature images along with the conventional PARS contrasts. Figure 5.9 (a-f) displays the members of $A$, along with their corresponding ground truth (Figure 5.9 (h)). It is worth mentioning that the three extracted features do exhibit a certain level of correlation, as shown in Figure 5.9 (g), however, the colorization results do show that these features are less redundant and better segment the structures compared to the conventional channels.

In the C-study, an exhaustive search is conducted on the set $A$, as explained earlier in this section. The objective is to determine the best combination of features for colorization. Given that there are six $(K + 3)$ elements in $A$ for the human skin data, there are a total of 63 possible combinations that can be used to train individual models. After training the 63 models, virtual staining is performed on unseen test data. This generates 63 colorized image sets using all the feature combinations from $A$. The test images are acquired from a distinct tissue section compared to the training data. These colorized test images are then compared against true H&E images.

Table 5.2 summarizes the quantitative assessment results. All of the employed metrics reached a consensus that there are at least 17 feature combinations that produce superior colorizations compared to the conventional PARS channels. Additionally, there are 13 combinations that outperform using all $N$ elements in $A$. This suggests that certain features may be redundant when combined and may potentially confuse the model learning process. Conversely, other features prove extremely valuable in colorization as they enhance the segmentation of tissue structures with distinct colors. This observation aligns with the hypothesis previously discussed in Section 5.1. Notably, the R channel is included in all of the top 30 feature combinations, highlighting the critical role that the R contrast plays in achieving precise colorization by offering distinctive and valuable biomolecule information.

The colorization results are depicted in Figure 5.10. The least satisfactory outcomes are obtained when using only one or two features, simply too limited to capture the complexity of different structures and their corresponding stained colors. An example using only $NR_{266}$ is shown in Figure 5.10 (b), which represents one of the poorest three combinations. With this very limited input information, the trained model appears to mistakenly identify RBCs as connective tissue and confuses the connective tissue with lipids, as highlighted in the yellow box. Conversely, using all the conventional channels ($NR_{266}$, $NR_{532}$, and $R_{266}$) in Figure 5.10 (c) demonstrates better performance, but still with shortcomings. For instance, some RBCs appear more purple than they do in the ground truth H&E, as highlighted in green. In addition, the colorization of the connective tissues sometimes exhibits a mixture of purple and pink shades instead of a consistent pink tone, as shown in the blue highlighted results. These artifacts may be attributed to insufficient input information, noise, or redundancy within the input data, all of which can hinder effective

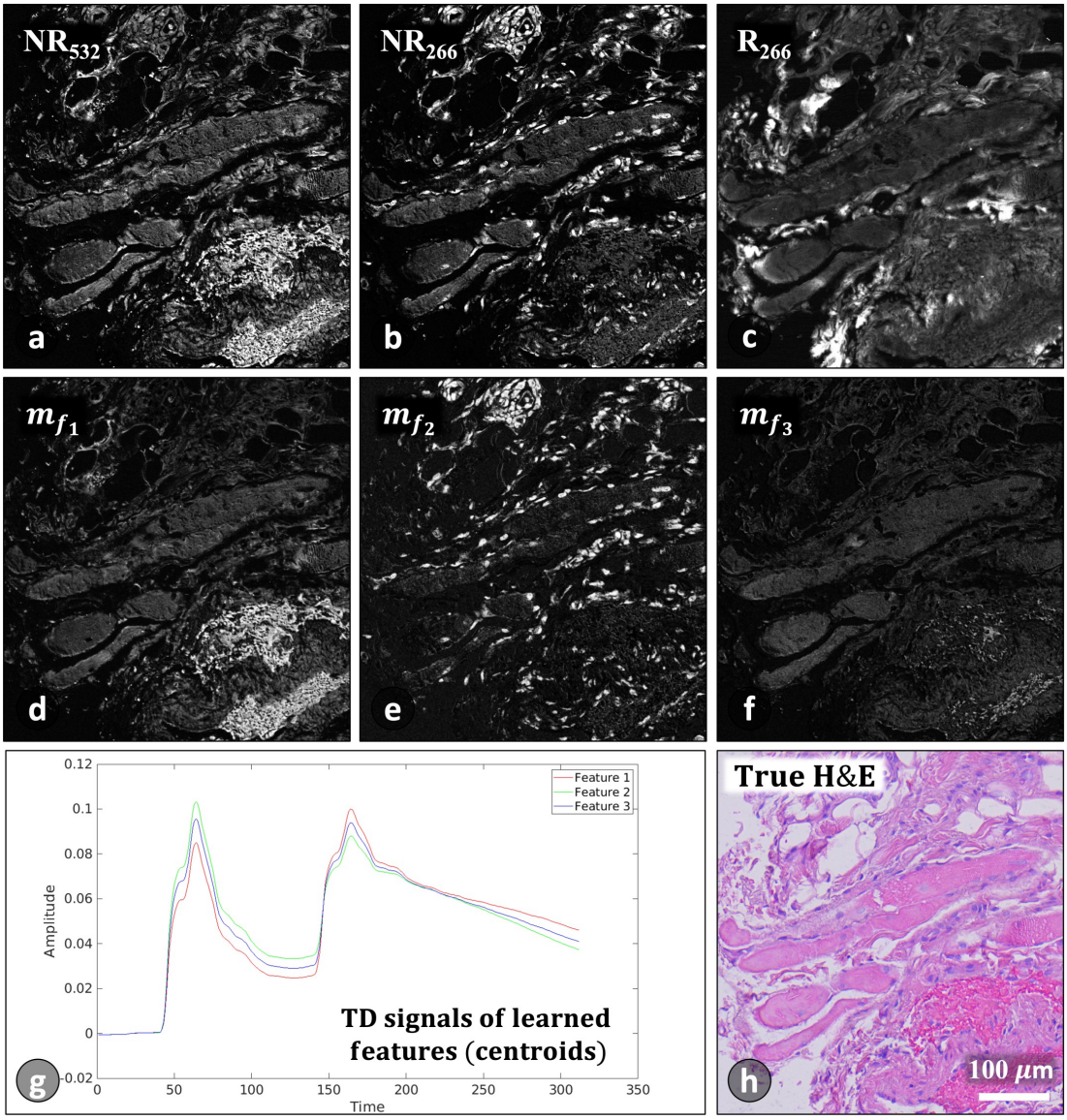



Figure 5.9: Example of PARS channels using human skin. (a) NR channel at 532 nm. (b) NR channel at 266 nm. (c) R channel at 266 nm. (d)-(f) Feature images ($m_{f_1}$, $m_{f_2}$ and $m_{f_3}$ corresponding to features 1-3 in (g), respectively. (g) TDs of 3 features extracted from NR channel. (e) True H&E of the same field-of-view.

Table 5.2: Summary of quantitative analysis of the C-study colorization results using the human skin dataset.

| Rank | Feature Combination | Score (rank) SSIM | PSNR | RMSE |
|---|---|---|---|---|
| Best | $NR_{532}$, $R_{266}$, $m_{f_1}$, $m_{f_3}$ | **0.89 (1)** | 22.92 (3) | 18.22 (3) |
| | $NR_{266}$, $R_{266}$, $m_{f_1}$ | 0.88 (2) | **23.03 (1)** | **17.99 (1)** |
| | $NR_{532}$, $NR_{266}$, $m_{f_2}$ | 0.88 (3) | 22.99 (2) | 18.06 (2) |
| ⋮ | | | | |
| Moderate | $NR_{532}$, $NR_{266}$, $R_{266}$, $m_{f_1}$, $m_{f_2}$, $m_{f_3}$ | 0.87 (14) | 22.50 (15) | 19.13 (15) |
| | $NR_{532}$, $NR_{266}$, $R_{266}$ | 0.87 (19) | 22.40 (18) | 19.35 (18) |
| ⋮ | | | | |
| Worst | $NR_{266}$ | 0.80 (61) | 19.92 (62) | 25.74 (62) |
| | $m_{f_2}$ | 0.72 (63) | 17.83 (63) | 32.71 (63) |

model learning.

In contrast, the feature combination that yielded the highest SSIM scores, and the top three RMSE and PSNR, is shown in Figure 5.10 (d). In the highlighted sections, colorization is the most accurate of the presented results. This is particularly prevalent in the RBCs and the shading of the connective tissue. These visual comparisons between the PARS virtual H&E and the true H&E images are supported by the quantitative measurements depicted in the leftmost column of Figure 5.10.

**Ablation Study.** Conducting a systematic ablation study is essential for gaining insights into the contribution and importance of individual features to the model's overall performance. By eliminating features one by one and retraining the model, this method enables the identification of critical features while distinguishing those that may carry lesser relevance. The results of this study, presented in Table 5.3, underscore the significance of each feature within the optimal feature combination. Notably, the $R_{266}$ channel emerges as the feature with the most substantial impact on performance, while the remaining features exhibit similar levels of influence.

In order to conduct a more detailed analysis of the outcomes from the C-study, we investigated specific results to assess the impact of the extracted features on the colorization output. Examples of these examinations are illustrated in Figure 5.11. Figure 5.11 (a) demonstrates the consequences of selecting features that do not predominantly represent

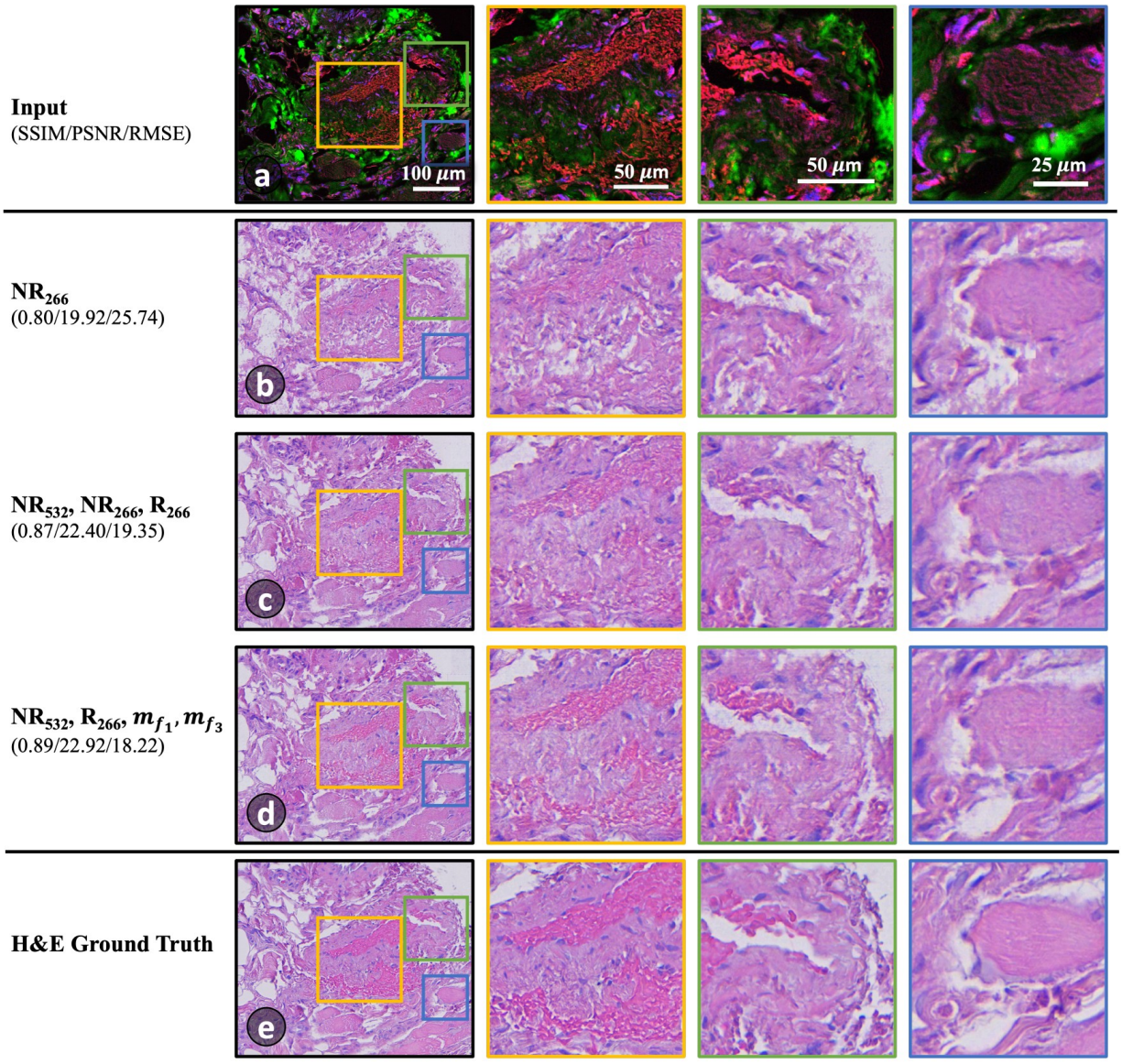



Figure 5.10: A comparison of virtual staining results of human skin using different combinations of PARS feature images as inputs. (a) RGB image of raw PARS data where R: $NR_{532}$, G: $R_{266}$, B: $NR_{266}$ (displayed for visualization), highlighting different parts of a human skin tissue sample. (b)-(d) Worst, moderate, and best results, respectively, using the feature combination labeled on the left. (e) True H&E image of the same field-of-view.

Table 5.3: Ablation study for the best feature combination, underscoring the significance of each individual feature on the model's performance.

| Feature Combination | Description | SSIM |
|---|---|---|
| $NR_{532}$, $R_{266}$, $m_{f_1}$, $m_{f_3}$ | Best Feature Combination | **0.886** |
| $R_{266}$, $m_{f_1}$, $m_{f_3}$ | (-$NR_{532}$) | 0.870 |
| $NR_{532}$, $m_{f_1}$, $m_{f_3}$ | (-$R_{266}$) | 0.838 |
| $NR_{532}$, $R_{266}$, $m_{f_3}$ | (-$m_{f_1}$) | 0.870 |
| $NR_{532}$, $R_{266}$, $m_{f_1}$ | (-$m_{f_3}$) | 0.876 |

cell nuclei. For a visual representation of the various structures emphasized in different feature images, refer to Figure 5.9. A notable degradation in the accurate colorization of nuclei, indicated by their incorrect association with connective tissue, is apparent when comparing the outcome to the ground truth in (f). This can be observed by the highlighted area with the green arrow.

Figure 5.11 (b) presents another example of such examinations. In this case, the features selected highlight both connective tissue and cell nuclei but do not highlight RBCs. As a result, the colorization of RBCs, highlighted with the red arrow in (f), appears noticeably faded and can be easily confused with the surrounding connective tissue (highlighted in yellow). Moreover, the colorization of the connective tissue itself appears pale. This observation highlights the crucial significance of the absence of specific structural information, as it can have a negative impact on the colorization not only of the missing structure but also on other structures within the image.

Subsequently, we examined the outcomes obtained when utilizing features that lacked information about two significant structures: connective tissue and RBCs. Specifically, we investigated the results obtained by utilizing two features that primarily focused on highlighting cell nuclei (Figure 5.11 (c)). This approach resulted in fairly accurate colorization solely for the cell nuclei, while the remaining structures in the image suffered from distortions in both color representation and structural integrity.

Finally, we aimed to compare the aforementioned three sets of results with the outcomes obtained by employing three features, each representing one of the major structures in the tissue: connective tissue, RBCs, and cell nuclei. Among all the results depicted in the figure, this specific outcome (shown in Figure 5.11 (d)) emerged as the most precise. It also ranked among the top three results of the entire C-Study. Remarkably, it visually resembled the ground truth (Figure 5.11 (f)) even more closely than the results obtained by utilizing all available features, depicted in Figure 5.11 (e). Additionally, the outcome in (d) is superior qualitatively to (e) as summarized in Table 5.2. This further validates the

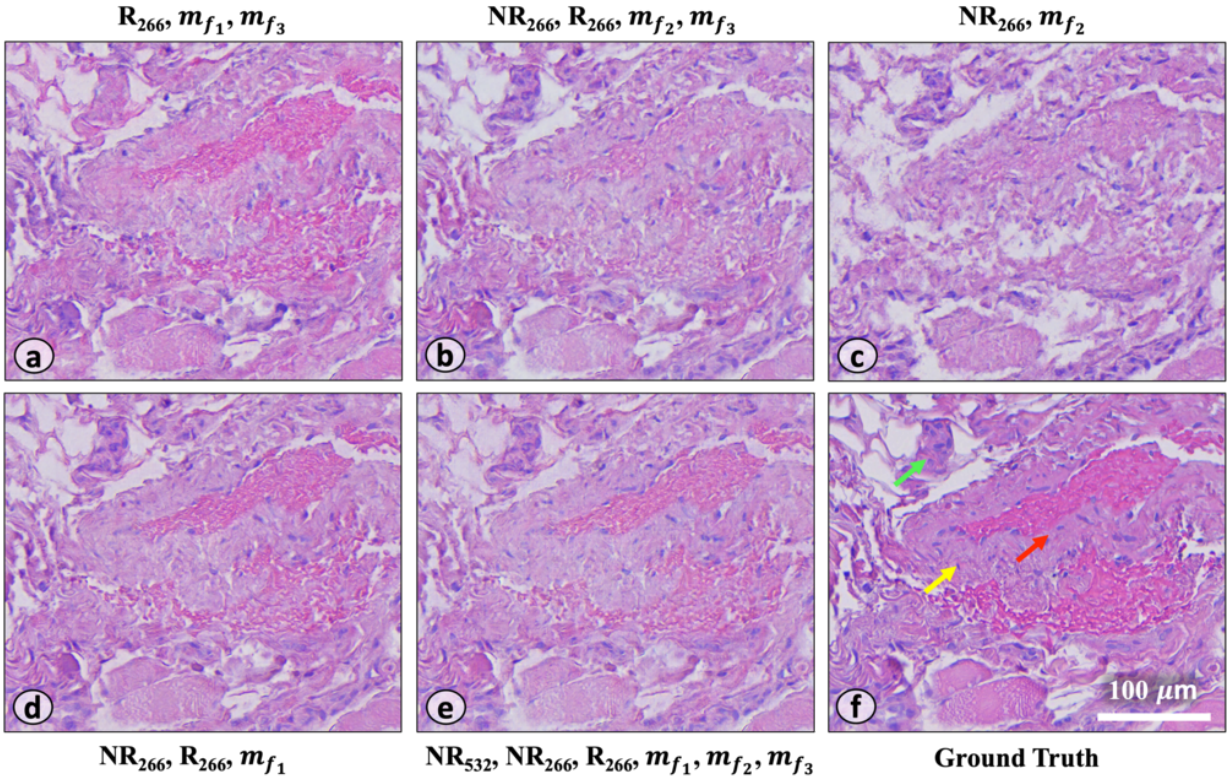



Figure 5.11: Further analysis of virtual staining results of human skin dataset using different combinations of PARS feature images as inputs. Three major structures are examined: connective tissue (yellow arrow), cell nuclei (green arrow), and RBCs (red arrow). (a)-(e) Colorization results using features labeled on each image as inputs. (f) True H&E ground truth.

notion that employing a greater number of features does not guarantee improved accuracy. Instead, it demonstrates that redundancy can potentially confuse the model, resulting in inferior outcomes.

The NR TD signals encompass information about the material properties being analyzed. Extracting features from these signals can significantly improve the labeling of biomolecules, leading to enhanced contrast. Notably, feature extraction surpasses the conventional image reconstruction methods in segmenting tissue structures such as nuclei, as presented in Figure 5.9 (b) and (e). The effectiveness of feature extraction in tissue labeling has been previously demonstrated in studies involving PARS data of murine brain

fresh tissue sample [94] and human breast tissue slide [113]. Incorporating these enhanced contrasts as input channels to a virtual staining model proves beneficial for colorization performance, ultimately enhancing the accuracy and visual representation of the colorized images.

Furthermore, the exhaustive search highlights that using an optimized set of features can be more advantageous than simply employing all available features. A targeted selection of features may lead to improved performance and reduced training time. This is likely because certain features may have a stronger ability to accurately label tissue structures, while others may be redundant and consequently cause confusion, negatively affecting the performance of the model.

### Mouse Brain Dataset

The mouse brain dataset shows a comparable performance pattern to the human skin. The workflow is replicated, starting with the K-study analysis. In this case, with an optimal $K$ value of 2, the number of channels in domain $A$ (PARS raw data) is 5 which allows for 31 trainable models with different channel input combinations. Figure 5.12 demonstrates that the results achieved when using $K = 2$ and $K = 3$ are similar both qualitatively and quantitatively. The figure also highlights the advantage of utilizing K$^*$-means features compared to the conventional NR channels. It is worth noting that, in this specific experiment, the number of epochs is set to be the same (200) across all the models with no early stopping criteria in order to evaluate the performance under identical parameters.

The results when $K = 2$ and $K = 3$ are very close both qualitatively and quantitatively as shown in Figure 5.12. The figure also illustrates the superiority of using K$^*$-means features in comparison with the conventional NR channels. Only in this experiment, the number of epochs is set to be the same to test the performance against the same exact parameters.

Following a comparison of all 31 input combinations, the quantitative assessment shows that there are at least 10 feature combinations that outperform the utilization of conventional features alone. Additionally, there are 8 alternative options that demonstrate superior performance compared to using all the features in domain $A$. Figure 5.13 presents a comparative analysis of the worst, moderate, and best colorization results achieved with various feature combinations as inputs to the model. The outcomes presented here further support the findings from the human dataset, which indicate that there are superior features for virtual staining that outperform the conventional features. Moreover, there

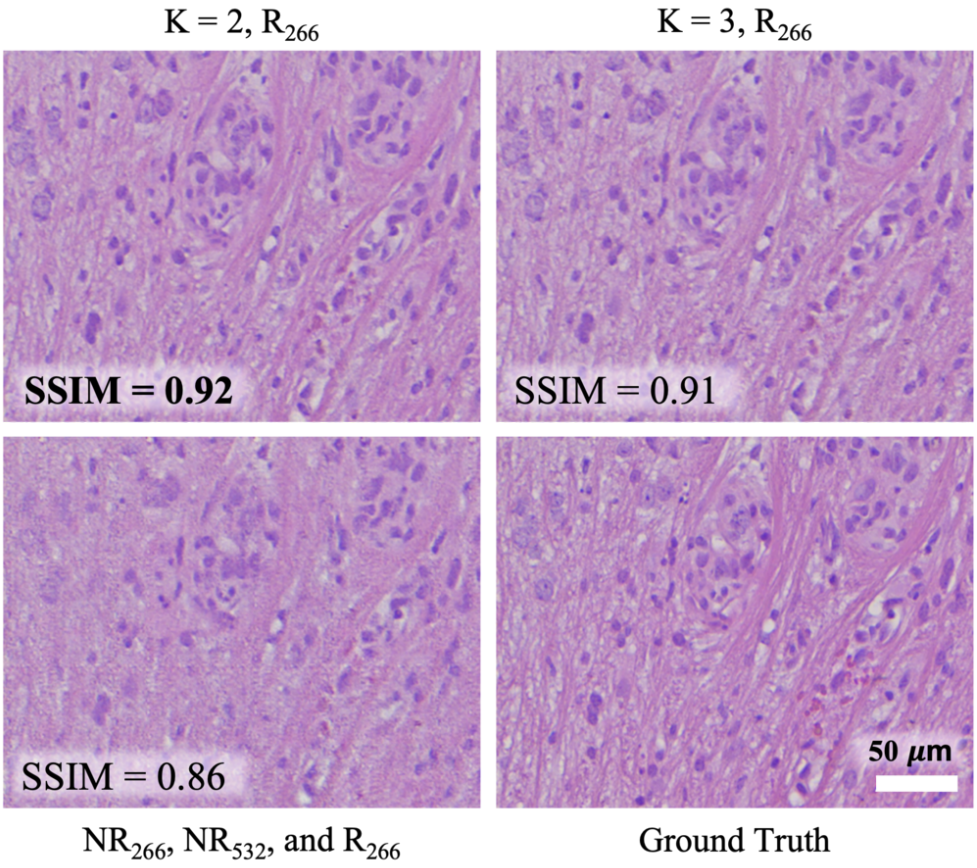



Figure 5.12: The virtual staining results comparing K*-means features (with $K = 2$ and $K = 3$) to the conventional NR channels. In all cases, $R_{266}$ is used as an additional input channel. The results demonstrate that using the extracted features at $K = 2$ and $K = 3$ are highly similar and outperform those using conventional channels.

is an optimal subset of features that produce the best results, as opposed to all features combined.

Figure 6 (b) illustrates the outcomes obtained from a feature combination ($R_{266}$ and $m_{f_1}$) that exhibit the poorest performance among the evaluated combinations. Notably, with this input combination, the trained model erroneously classifies parts of connective tissue as cell nuclei. Upon observing Figure 5.13 (b) and (c), it becomes apparent that connective tissue structures and tissue surrounding the tumor, which should ideally exhibit a distinct pink color (as demonstrated in the ground truth image, (e)), are more accurately colorized in (d) compared to (c). In (c), a significant portion of the pink color is substituted with shades of brown. The model can sometimes employ an averaging strategy for colors as a means to minimize loss [172], resulting in the prediction of colors like gray or brown when uncertain about the best color choice. In contrast, the highest level of accuracy among the presented results is observed in (d), which exhibits significant correspondence with the ground truth in terms of visual appearance and quantitative measurements.

These observations provide additional evidence that the extracted features contribute to the learning process of the model. Furthermore, they emphasize the significance of selecting features that accurately label and represent the data, ultimately resulting in improved overall performance. The presented results highlight that superior feature extraction methods and feature combinations exist for improved colorization of the raw input data. Future efforts will explore other tissue types and staining varieties.

## 5.10 Conclusion

In conclusion, this chapter explores the use of time-resolved non-radiative signals for improving the virtual staining of PARS images in both human skin and mouse brain. Using the modified K-means method [94, 113] to extract features from the non-radiative TD signals, additional information about imaged targets is captured. The proposed MC-GAN extends the conventional colorization model to accept more than three channels, allowing for the utilization of these additional features. The experimental results demonstrate that certain feature combinations outperform the conventional PARS channels as they exhibit improved labeling of tissue structures.

Several experiments are conducted to determine the optimum number of K*-means features ($K$) as well as the optimum feature combination for training virtual staining models. Three different metrics are employed to evaluate the model's performance for feature selection. The limitations of using only one or two features are evident, as they fail

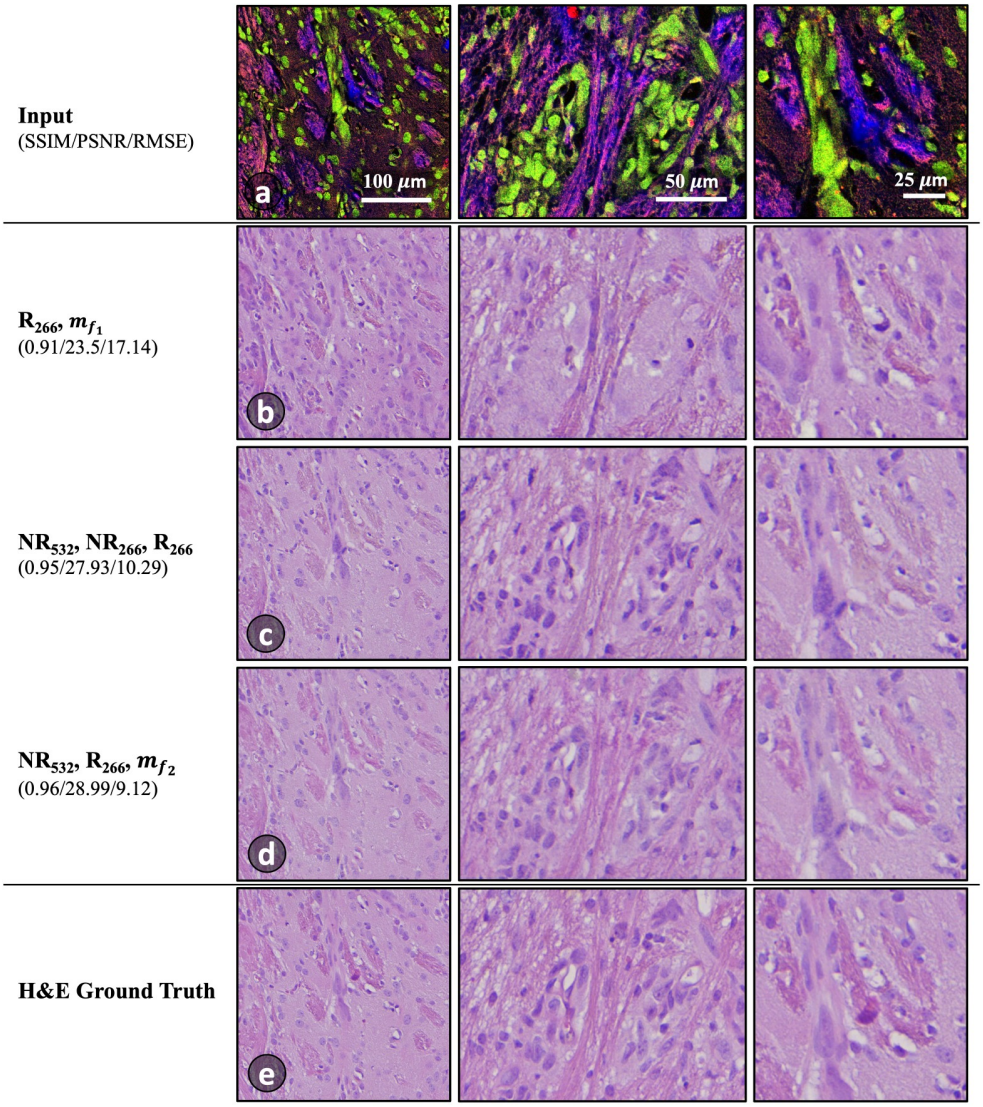



Figure 5.13: Another comparison of virtual staining results of mouse brain tissue. (a) RGB image of raw PARS data where R: $NR_{532}$, G: $R_{266}$, B: $NR_{266}$ (displayed for visualization). (b)-(d) Worst, moderate, and best results, respectively, using the feature combination labeled on the left. (e) True H&E image of the same field-of-view.

to accurately represent the complexity of different structures and their colors. Moreover, employing all the available features can lead to confusion within the model due to the potential redundancy among them. Therefore, it is crucial to conduct a comprehensive search to identify the most effective feature combinations, which not only reduced training time by utilizing fewer features but also alleviated model confusion.

With the optimal feature combination, the colorization results exhibit a high degree of accuracy, as observed in the results. These findings highlight both high visual and quantitative agreement between the H&E-like PARS and the true H&E images among the two datasets, emphasizing the potential of TD signals in enhancing the accuracy of virtual staining techniques.

# Chapter 6

# Application: PARS-to-PAS Translation of Human Skin Fungus

## 6.1 Introduction

In previous works [1,2,97,164], the application of PARS data has shown remarkable efficacy in generating virtual H&E stains. This breakthrough has sparked interest in exploring the potential of PARS technology with other stains, as it offers the possibility of applying multiple stains to a single tissue section using deep learning techniques. This advancement holds great promise for improving the accuracy and consistency of tissue-based diagnoses, allowing pathologists to extract more relevant information from tissue samples than was previously feasible [173].

Among the various stains used in histopathology, the PAS stain (presented in Section 2.1.2) holds significant importance, particularly in the evaluation of certain diseases, including those caused by fungal infections [36, 174]. PAS staining facilitates the identification and characterization of fungal structures within tissue sections. Given its clinical relevance, there is a need to investigate the translation of PARS data into virtual PAS stains, enabling pathologists to leverage the benefits of PARS technology for the diagnosis of various conditions.

In this chapter, we focus on experimenting with PARS data, both in its raw form and by extracting relevant features as detailed in Chapter 5, to generate virtual PAS stains. Our work in this chapter specifically targets human skin fungus tissue sections.

The exploration of PARS data in generating virtual stains beyond H&E marks a novel

advancement, paving the way for enhancing histopathological analysis. By incorporating deep learning techniques and integrating extracted features from PARS data, there is considerable potential to improve the visualization and interpretation of diverse tissue structures. This promising approach can offer valuable insights to support pathologists in achieving precise diagnoses and making informed decisions regarding subsequent treatment options.

## 6.2 Dataset Preparation

The dataset used in this chapter is obtained from thin sections of formalin-fixed paraffin-embedded (FFPE) human skin tissues. The samples were provided by clinical collaborators and collected under approved protocols by the Research Ethics Board of Alberta and the University of Waterloo Health Research Ethics Committee. Patient consent was waived as the samples were archival tissues with no patient identification. The experiments followed the ethical guidelines set by the government of Canada.

The unstained tissue sections underwent fixation, dehydration, and paraffin embedding. Thin sections of approximately 5 $\mu m$ thickness were sliced and affixed to glass slides, and excess paraffin was removed through baking.

PARS image preprocessing and registration follow the same method presented in Chapters 3 and 5. To collect the dataset for virtual staining, unstained tissue sections were imaged using the single-excitation-wavelength PARS system described in Section 3.3. The same tissue sections were then stained using H&E and scanned with a transmission-mode bright-field microscope at 40x magnification (Morpholens 1, Morphle Digital Pathology). Subsequently, the same tissue section underwent destaining with 1% acid alcohol for a duration of 1 minute, followed by a rinse with xylene for 1 minute. It was then stained using PAS stain. By following this procedure, we obtain raw PARS, H&E, and PAS images of the exact same sample. It is important to note that the virtual staining model requires training with the ground truth data generated from the standard staining procedure. However, once the model is trained, the staining step is no longer necessary.

## 6.3 Training Settings

The cycleGAN model described in Section 5.7 is employed to translate the label-free PARS images into virtually stained images.

To create the training sets for the human skin data, overlapping patches of size $256 \times 256$ pixels were extracted from the PARS, H&E, and PAS images using one sample. These pairs were divided, with 70% used for training, 10% for validation, and 20% for testing purposes. Following the same values from Section 4.2.4, the learning rate was set to 0.0002, and the training process continued for a maximum of 200 epochs. Early stopping was implemented, terminating the training when the generator loss no longer improved. Subsequently, the trained model is applied to unseen test images, which were also divided into overlapping patches of size $256 \times 256$ pixels. To ensure a smooth transition between adjacent patches, a practical approach involved employing an overlap of approximately 50%. This method is found to effectively reduce any noticeable artifacts along the borders of the final stitched image, as determined through empirical experimentation.

## 6.4 Results and Discussion

The primary objective of this chapter is to translate our PARS data into PAS visualizations. To achieve this, we conducted several experiments. First, we trained a model using conventional PARS channels ($NR_{266}$ and $R_{266}$) along with their corresponding PAS ground truth data. For this experiment, we extracted about 2000 overlapping patches of size $256 \times 256$. The preprocessing and registration procedures were performed following the same approaches described in earlier chapters.

The outcomes of this experiment are depicted in Figure 6.1, which presents a comprehensive view of human skin fungus tissue, including both large-field-of-view and zoomed-in figures. The findings exhibit a strong visual agreement with the ground truth data in multiple regions, such as the fungal structures presented in magenta color and highlighted by the red box, as well as nuclei and connective tissue highlighted by the blue box. However, it is worth noting that the colorization of the membrane surrounding the sweat gland structures (highlighted in green) is not accurately rendered, as it should be magenta. This discrepancy may be attributed to the limited number of examples of these specific structures within the training dataset.

In addition, our second objective is to directly compare the visualizations H&E, and PAS using PARS input data. Each stain offers complementary information, with PAS highlighting fungi and specific carbohydrates, while H&E emphasized nuclei and provided a broader view of tissue morphology, as presented in Section 2.1.1. The unified approach aims to gain a comprehensive understanding of the tissue's composition, potentially aiding in accurate diagnosis.

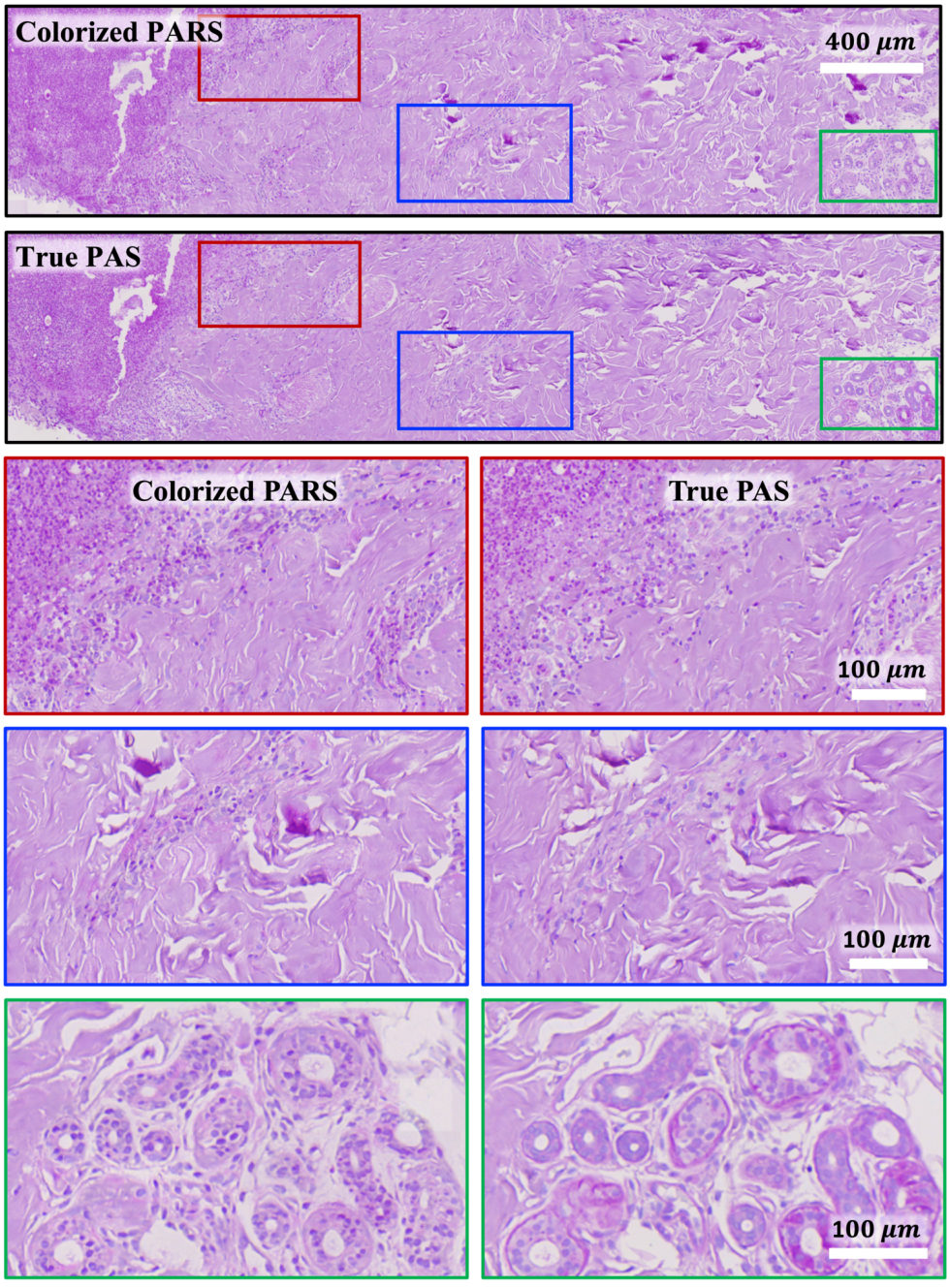



Figure 6.1: Virtual PAS staining of PARS images using conventional channels.

Moreover, as demonstrated in the preceding chapter, incorporating additional features consistently yields better results compared to using conventional data alone. Therefore, we employed the multi-channel framework introduced in Chapter 5 to expand the input channels by utilizing features extracted from the NR TD signals.

The human skin dataset used in this case involves a single excitation wavelength of 266 nm. By incorporating the extracted features, we enhance the information available at each pixel location, thereby improving the colorization model's ability to understand the statistical transformation between the source and target domains, as discussed in the previous chapter.

Additionally, the results from the previous chapter demonstrated that employing $K = 2$ or $K = 3$ for feature extraction consistently outperforms other values of $K$ by a considerable margin. Consequently, we decided to conduct our experiments solely with these two values for $K$. While these values may not represent the optimal choices for $K$ for this specific dataset, insights from past experiments suggest that utilizing these specific values of $K$ for feature extraction and combining the resultant features with conventional channels yields better performance than using conventional channels alone. It is important to note that this specific dataset differs from the one used in the previous chapter in terms of the number of excitation wavelengths utilized for PARS data and the type of stains employed for the ground truth data.

Considering the preliminary and exploratory nature of the work presented in this chapter, intended to serve as a proof of concept, we made a deliberate choice to adopt a shortcut strategy by selecting specific combinations likely to yield enhancements. Nevertheless, it is essential to acknowledge that achieving optimal results would require conducting a comprehensive search, similar to the methodology presented in Chapter 5.

Consequently, six different models were trained for the second experiment:
(1) PARS to H&E using conventional PARS data.
(2) PARS to H&E using $NR_{266}$ features extracted with $K = 2$ in addition to the conventional channels.
(3) PARS to H&E using $NR_{266}$ features extracted with $K = 3$ in addition to the conventional channels.
(4) PARS to PAS using conventional PARS data.
(5) PARS to PAS using $NR_{266}$ features extracted with $K = 2$ in addition to the conventional channels.
(6) PARS to PAS using $NR_{266}$ features extracted with $K = 3$ in addition to the conventional channels.

The six models were trained using approximately 300 patches. The extracted patches

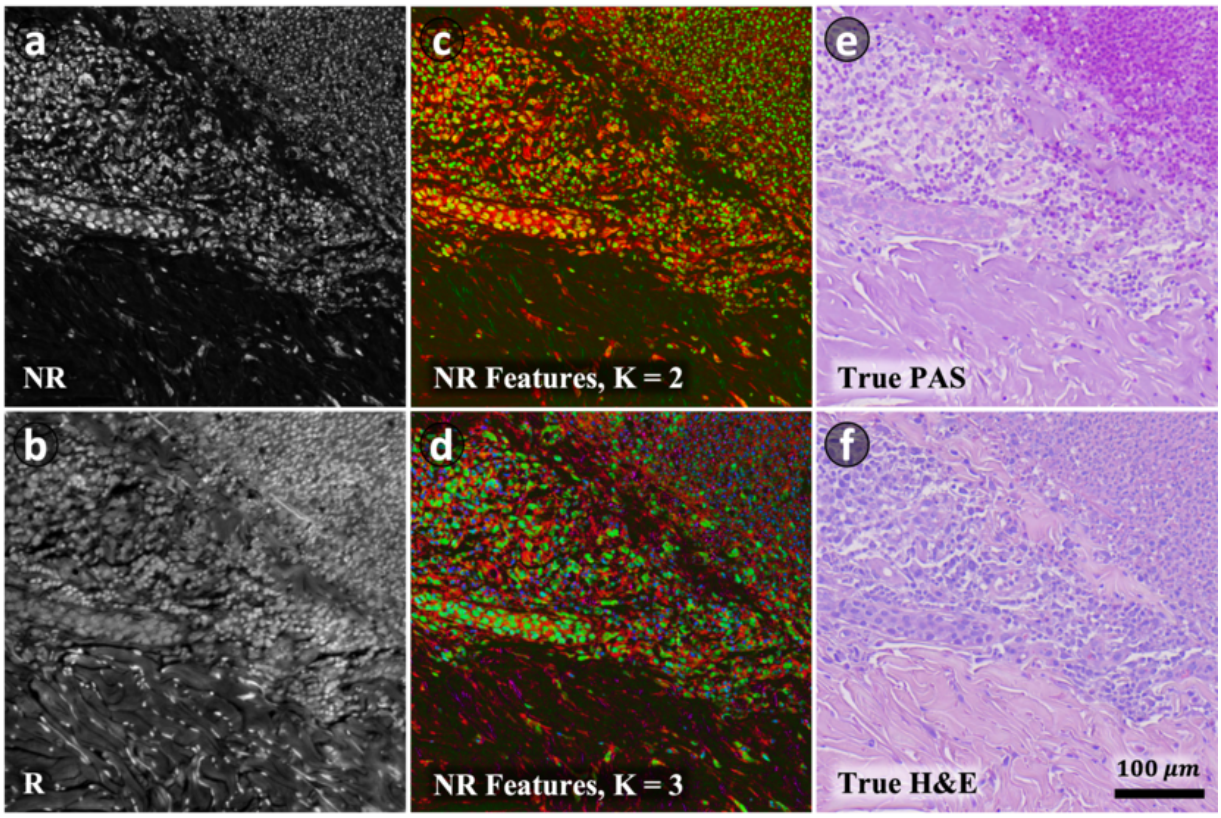



Figure 6.2: An illustration of various image domains of the same human skin tissue section. (a), (b) Raw PARS conventional channels (NR and R channels, respectively). (c), (d) PARS K-means features extracted from NR TD signals using $K = 2$ and $K = 3$, respectively combined as RGB images. (e), (f) Chemically stained PAS and H&E images, respectively.

were the same across all image domains to ensure a fair comparison. Figure 6.2 illustrates an example of the different inputs employed in this experiment along with their corresponding true H&E and PAS counterparts from the same section.

Figure 6.2 displays a field of view of the tissue that encompasses cell nuclei, fungal structures, and extracellular structures. The PARS NR image (Figure 6.2 (a)) primarily highlights nuclei, as the excitation wavelength used for this dataset is 266 nm, which predominantly targets DNA. On the other hand, the R channel image (Figure 6.2 (b)) emphasizes the majority of the other structures.

In Figure 6.2 (c) and (d), we can observe the RGB images of the extracted features when

Table 6.1: Quantitative analysis summary of H&E and PAS colorization results on human skin dataset, considering three scenarios: first, using only conventional channels; second, incorporating conventional channels alongside extracted features when $K = 2$; and third, when $K = 3$.

| Stain | Input Channels | SSIM |
|---|---|---|
| H&E | Conventional only | 0.85 |
| | Conventional + Extracted (K=2) | 0.86 |
| | Conventional + Extracted (K=3) | **0.88** |
| PAS | Conventional only | 0.81 |
| | Conventional + Extracted (K=2) | **0.83** |
| | Conventional + Extracted (K=3) | 0.79 |

$K = 2$ and $K = 3$, respectively. When $K = 2$, the structures present in the NR image were plausibly divided into three primary features: extracellular structures mostly represented in various shades of red, fungal structures represented in green, and cell nuclei represented in yellow (a combination of the two other feature shades). The yellow color indicates that cell nuclei were not extracted as an independent feature; instead, they comprise a portion from each of the other two features.

On the other hand, when $K = 3$ (Figure 6.2 (d)), cell nuclei are manifested as a distinct feature colored in green, fungal structures are generally colored in blue, and extracellular structures remain colored in red. While there is no direct evidence of the extracted features correlating with biological structures, the visualizations presented reinforce the idea that the employed feature extraction method demonstrates superior proficiency in segmenting and labeling structures compared to conventional channels, potentially yielding more effective results.

The evaluation of the results (summarized in Table 6.1) involved calculating the SSIM between the colorized images and their corresponding ground truth. In the case of H&E models, using $K = 3$ achieved a 1.15% higher score compared to $K = 2$ and a 2.77% improvement over conventional channels. On the other hand, for PAS models, using $K = 2$ resulted in a 4.56% better performance than $K = 3$ and a 2.35% improvement over using conventional channels. The results can be observed in Figure 6.3. Once again, the effectiveness of employing additional extracted channels is demonstrated not only with H&E stains but also with other stains like PAS.

The results indicate that it is feasible to generate multiple stains from the same input PARS data, which is valuable for diagnosis [173]. This approach allows for preserving the tissue for further analysis instead of resorting to restaining. Furthermore, the experi-

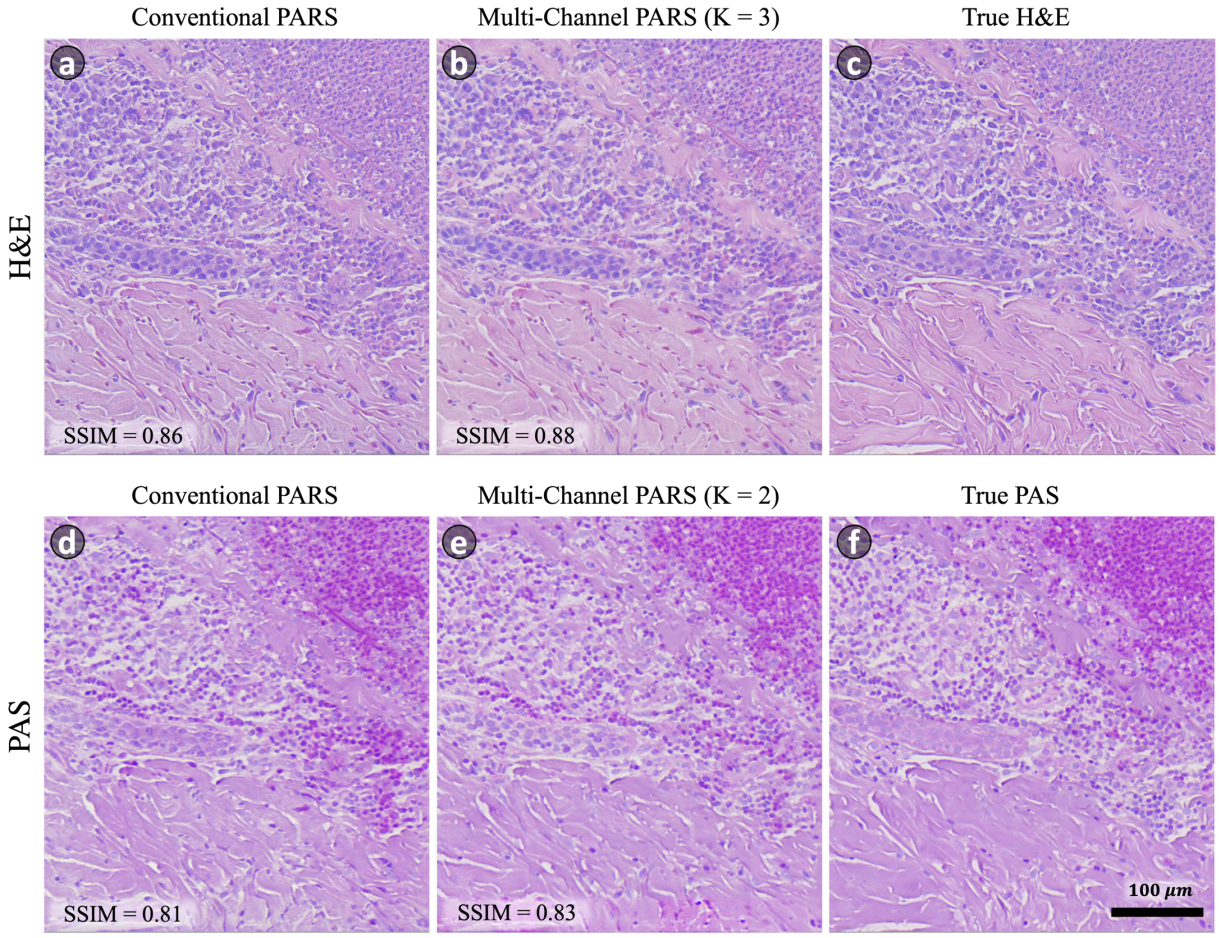



Figure 6.3: A comparison between virtual Staining results using conventional channels alone and using additional best performing $K$ channels for H&E ($K = 3$) and PAS ($K = 2$) stains.

mental results demonstrate that even when the tissue type and dataset are identical, the optimal values for $K$ vary for each stain. This implies that different stains require different information to achieve the best results.

Additional investigations were conducted to assess the efficacy of employing multiple channels. Figure 6.4 presents a few examples. For instance, in Figure 6.4 (a), it is evident that when using multi-channels instead of conventional channels alone, the nuclei with their outlines exhibit substantially enhanced contrast and visibility, even though this improvement might not be readily apparent in the ground truth data.

In Figure 6.4 (b), the majority of the nuclei that were lost when using conventional channels are effectively restored when multiple channels are employed. Additionally, with conventional data, there is confusion between fungal structures and nuclei, which was rectified when utilizing multi-channel input. Finally, certain connective tissue structures shown in Figure 6.4 (c) are misidentified as fungal structures, resulting in some sections being inaccurately colored in magenta instead of the expected shades of pink.

These findings offer supplementary evidence supporting the notion that the extracted features play a crucial role in enhancing the model's learning process. They underscore the importance of accurate labeling and representation of the data, leading to overall improved performance. The results presented here emphasize the existence of effective feature extraction methods, which ultimately contribute to enhanced colorization of the input data.

Based on these valuable insights, future research endeavors will expand upon the current exploratory work to encompass larger datasets and incorporate different tissue types. Additionally, the investigation will extend to encompass various staining techniques. The objective is to analyze how each staining method impacts the optimal choice of the parameter $K$ for feature extraction.

By exploring bigger datasets and diverse tissue types, researchers can gain a more comprehensive understanding of the colorization process and its adaptability across various biological structures. The comparison of different staining techniques will provide valuable information on how the choice of stain affects the effectiveness of feature extraction and the subsequent colorization results.

This extended research will not only contribute to advancing the field of image analysis and colorization but also facilitate the development of more robust and accurate algorithms for various biomedical applications. Ultimately, it will empower scientists and researchers to gain deeper insights into the complexities of biological structures and enhance their ability to interpret and analyze medical imaging data.

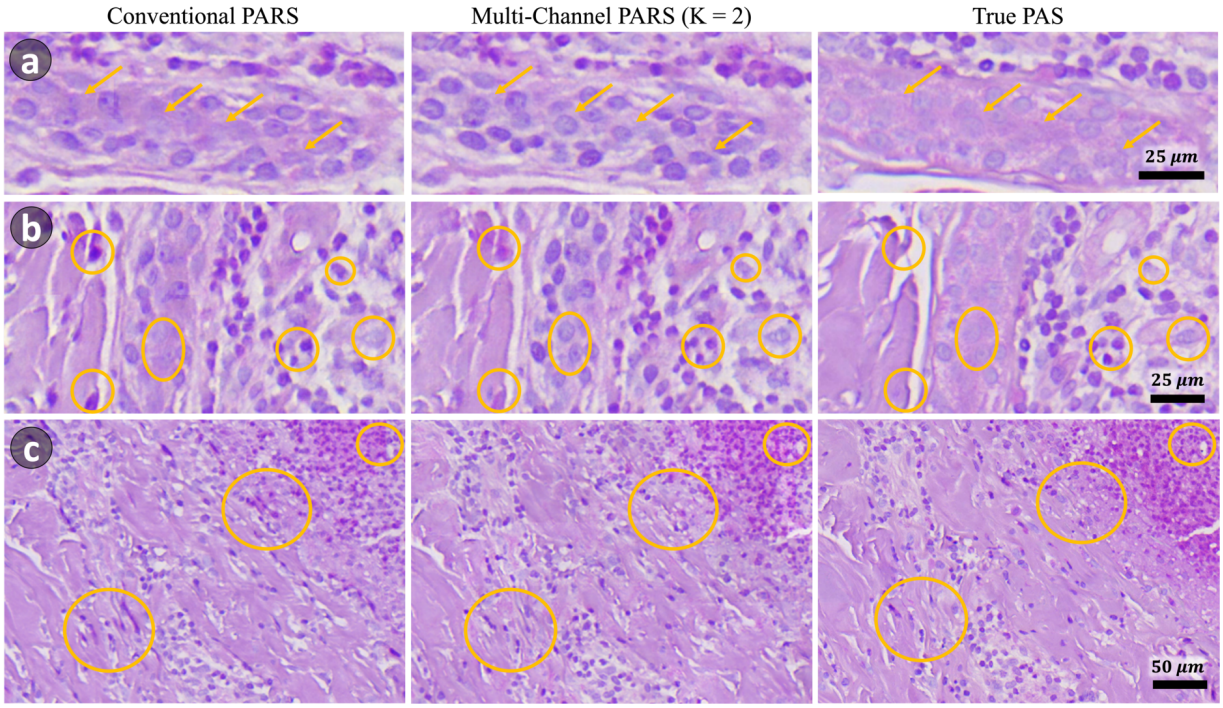



Figure 6.4: Further examination of virtual PAS staining using both conventional and multiple channels as inputs to the model. (a) Nuclei with enhanced contrast are visible, demonstrating the advantage of multi-channel over conventional PARS. (b) Most lost nuclei from conventional channels are restored, and confusion between fungal structures and nuclei is mostly resolved. (c) Misidentification of connective tissue structures as fungal structures is corrected when utilizing multi-channel input.

## 6.5 Conclusion

To summarize, this research investigates the application of conventional PARS channels and NR TD signals in generating PAS stains. By employing the K-means method, as described in Chapter 5, feature extraction from NR TD signals allows for capturing additional information about the imaged targets. The experimental findings indicate that the utilization of these additional features surpasses the performance of conventional PARS channels in both H&E and PAS stains. This approach exhibits enhanced labeling of tissue structures, highlighting its potential for improving different staining techniques. Furthermore, this approach demonstrates the potential of PARS to generate virtual stains using the same data sources.

# Chapter 7

# Conclusions and Future Work

In conclusion, this thesis has made significant progress in developing virtual histological staining using label-free PARS microscopy. The successful application of state-of-the-art deep learning models such as pix2pix (described in Chapter 3) in replicating H&E stains from PARS images for the first time, combined with the confirmation of diagnostic quality by pathologists, highlights their potential as an efficient diagnostic tool.

Chapter 4 focused on the comparison of paired (pix2pix) and unpaired (cycleGAN) deep learning-based virtual staining models and the impact of aligning PARS and H&E images on their performance. Our findings revealed that each model has its strengths and weaknesses. Pix2pix produced effective digital H&Es but necessitated precise image alignment. On the other hand, cycleGAN handled unpaired data but exhibited possible inaccuracies. This research highlights the ongoing need for improvements to ensure reliable diagnostic applications.

The core objective of this thesis was to advance the field of virtual histology by exploring the capabilities of PARS microscopy with deep learning algorithms. The novel multi-feature virtual staining pipeline, introduced in Chapter 5, incorporating PARS time-resolved signals and feature selection methods, significantly improved structure distinction and tissue labeling, leading to enhanced colorization process. Furthermore, the pioneering translation of PARS data into PAS stain (outlined in Chapter 6) expands the versatility of the proposed frameworks, presenting new opportunities for research and diagnostics with various staining types. Further validation and comparison with traditional staining methods are crucial to ensure the widespread adoption and reliability of virtual staining of PARS histology in diverse clinical scenarios.

Looking ahead, we can identify various key pathways for further progress. One crucial

step is to invest resources in building a large-scale dataset encompassing samples of different tissue types and stains. A comprehensive and diverse dataset will be essential for thorough validation and eventual clinical adoption. This well-curated dataset will enhance the generalization of deep learning models, improving their performance on a wide range of histological staining tasks. Moreover, leveraging the power of large-scale datasets, it becomes feasible to create pre-trained models tailored for the virtual staining of PARS images. By training these models on such datasets, they can serve as valuable initial weights that can be fine-tuned for specific tasks like classification, allowing for efficient transfer learning in various histological analyses.

Developing automatic registration tools explicitly tailored for PARS data is necessary to address registration issues and improve the performance of the pix2pix model. Overcoming blurriness and ensuring accurate alignment will enhance the quality of colorized images and the overall performance. Lastly, bridging the gap between pix2pix and cycleGAN offers an exciting prospect for developing hybrid approaches that combine the strengths of both models. For instance, researchers have employed an alternating training strategy to bridge this gap [175]. This approach allowed cycleGAN to efficiently learn explicit pixel-level mappings from paired data while simultaneously learning implicit domain mappings from unpaired data. To handle the mixed data, various pair consistency losses were introduced for paired and unpaired data, enhancing the overall performance of the hybrid approach.

To further enhance colorization accuracy and overall performance, experimentations with different generator architectures for GAN models can be explored. Typical GAN models use the U-Net model as a generator which has proven effective, however, alternative generators may offer additional improvements in colorization accuracy. Another promising avenue for improvement lies in enhancing feature extraction methods. By exploring and developing feature extraction techniques, more information per pixel location can be obtained, leading to better labeling of tissue structures and thus improved colorization accuracy.

Validation efforts should be expanded to encompass more stains and datasets in order to ensure the robustness and reliability of virtual histology. A wider range of stains and data will provide a comprehensive assessment of the capabilities of PARS virtual staining and its potential limitations. Moreover, exploring whole-slide colorization can yield valuable insights for real-world applications and large-scale tissue analysis. Finally, incorporating clinical validation is a crucial step toward the clinical adoption of virtual histological staining. Collaborating with clinical experts and involving them in the validation process will ensure that virtual staining meets the necessary standards for medical diagnosis.

# References


[1] M. Boktor, B. R. Ecclestone, V. Pekar, D. Dinakaran, J. R. Mackey, P. Fieguth, and P. Haji Reza, "Virtual histological staining of label-free total absorption photoacoustic remote sensing (TA-PARS)," *Scientific Reports*, vol. 12, p. 10296, June 2022.

[2] M. Boktor, J. E. D. Tweel, B. R. Ecclestone, J. A. Ye, P. Fieguth, and P. H. Reza, "Multi-channel feature extraction for virtual histological staining of photon absorption remote sensing images," 2023.

[3] J. Yao and L. V. Wang, "Sensitivity of photoacoustic microscopy," *Photoacoustics*, vol. 2, pp. 87–101, June 2014.

[4] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, pp. 53–65, jan 2018.

[5] Y. Winetraub, E. Yuan, I. Terem, C. Yu, W. Chan, H. Do, S. Shevidi, M. Mao, J. Yu, M. Hong, and et al. *OCT2Hist: Non-invasive virtual biopsy using optical coherence tomography*, 2021.

[6] V. Baxi, R. Edwards, M. Montalto, and S. Saha, "Digital pathology and artificial intelligence in translational medicine and clinical practice," *Modern Pathology*, vol. 35, no. 1, pp. 23–32, 2022.

[7] C. E. Day, ed., *Histopathology: Methods and Protocols*, vol. 1180 of *Methods in Molecular Biology*. New York, NY: Springer New York, 2014.

[8] L. Kang, X. Li, Y. Zhang, and T. T. Wong, "Deep learning enables ultraviolet photoacoustic microscopy based histological imaging with near real-time virtual staining," *Photoacoustics*, vol. 25, p. 100308, 2022.

[9] J. Li, J. Garfinkel, X. Zhang, D. Wu, Y. Zhang, K. de Haan, H. Wang, T. Liu, B. Bai, Y. Rivenson, G. Rubinstein, P. O. Scumpia, and A. Ozcan, "Biopsy-free in vivo virtual histology of skin using deep learning," *Light: Science & Applications*, vol. 10, p. 233, Dec. 2021.

[10] S. Aumann, S. Donner, J. Fischer, and F. Müller, "Optical Coherence Tomography (OCT): Principle and Technical Realization," in *High Resolution Imaging in Microscopy and Ophthalmology: New Frontiers in Biomedical Optics* (J. F. Bille, ed.), pp. 59–85, Cham: Springer International Publishing, 2019.

[11] D. G. Eichberg, A. H. Shah, L. Di, A. M. Semonche, G. Jimsheleishvili, E. M. Luther, C. A. Sarkiss, A. D. Levi, S. H. Gultekin, R. J. Komotar, and M. E. Ivan, "Stimulated Raman histology for rapid and accurate intraoperative diagnosis of CNS tumors: prospective blinded study," *Journal of Neurosurgery*, vol. 134, pp. 137–143, Jan. 2021.

[12] B. Sarri, F. Poizat, S. Heuke, J. Wojak, F. Franchi, F. Caillol, M. Giovannini, and H. Rigneault, "Stimulated Raman histology: one to one comparison with standard hematoxylin and eosin staining," *Biomedical Optics Express*, vol. 10, p. 5378, Oct. 2019.

[13] Y. Rivenson, H. Wang, Z. Wei, K. Haan, Y. Zhang, Y. Wu, H. Gunaydin, J. Zuckerman, T. Chong, A. Sisk, L. Westbrook, W. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nature Biomedical Engineering*, vol. 3, June 2019.

[14] B. Bai, X. Yang, Y. Li, Y. Zhang, N. Pillar, and A. Ozcan, "Deep learning-enabled virtual histological staining of biological samples," *Light: Science & Applications*, vol. 12, p. 57, Mar. 2023.

[15] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," *arXiv:1406.2661 [cs, stat]*, June 2014. arXiv: 1406.2661.

[16] P. Hajireza, W. Shi, K. Bell, R. J. Paproski, and R. J. Zemp, "Non-interferometric photoacoustic remote sensing microscopy," *Light: Science & Applications*, vol. 6, pp. e16278–e16278, June 2017.

[17] B. R. Ecclestone, K. Bell, S. Sparkes, D. Dinakaran, J. R. Mackey, and P. H. Reza, "Label-free virtual Hematoxylin and Eosin (H&E) staining using second generation Photoacoustic Remote Sensing (PARS)," p. 38.

[18] 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)*, (Honolulu, HI), pp. 5967–5976, IEEE, July 2017.

[19] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, Oct. 2017. ISSN: 2380-7504.

[20] "Periodic Acid Schiff (PAS) Staining: A Useful Technique for Demonstration of Carbohydrates," *Medico-Legal Update*, Apr. 2020.

[21] A. P. Brown, P. Drew, B. Knight, P. Marc, S. Troth, K. Wuersch, and J. Zandee, "Graphical display of histopathology data from toxicology studies for drug discovery and development: An industry perspective," *Regulatory Toxicology and Pharmacology*, vol. 82, pp. 167–172, 2016.

[22] M. Y. Zaidi, R. Canter, and K. Cardona, "Post-operative surveillance in retroperitoneal soft tissue sarcoma: The importance of tumor histology in guiding strategy," *Journal of Surgical Oncology*, vol. 117, no. 1, pp. 99–104, 2018.

[23] A. H. Fischer, K. A. Jacobson, J. Rose, and R. Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold Spring Harbor Protocols*, vol. 2008, no. 5, 2008.

[24] M. El-Bahrawy and R. Ganesan, "Frozen section in gynaecology: Uses and limitations," *Archives of Gynecology and Obstetrics*, vol. 289, no. 6, p. 1165–1170, 2014.

[25] H. Jaafar, "Intra-Operative Frozen Section Consultation: Concepts, Applications and Limitations," *The Malaysian Journal of Medical Sciences : MJMS*, vol. 13, pp. 4–12, Jan. 2006.

[26] H. A. Alturkistani, F. M. Tashkandi, and Z. M. Mohammedsaleh, "Histological Stains: A Literature Review and Case Study," *Global Journal of Health Science*, vol. 8, p. 72, June 2015.

[27] J. K. Chan, "The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology," *International Journal of Surgical Pathology*, vol. 22, no. 1, p. 12–32, 2014.

[28] C. Ortiz-Hidalgo and S. Pina-Oviedo, "Hematoxylin: Mesoamerica's gift to histopathology. palo de campeche (logwood tree), pirates' most desired treasure,

and irreplaceable tissue stain," *International Journal of Surgical Pathology*, vol. 27, no. 1, p. 4–14, 2018.

[29] S. K. Suvarna, *Bancroft's theory and practice of histological techniques consult.* Churchill Livingstone, Elsevier, 2013.

[30] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[31] Y. Li, N. Li, X. Yu, K. Huang, T. Zheng, X. Cheng, S. Zeng, and X. Liu, "Hematoxylin and eosin staining of intact tissues via delipidation and ultrasound," *Scientific Reports*, vol. 8, no. 1, 2018.

[32] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[33] J. C. E. Underwood, "The Use of Stains," in *Introduction to Biopsy Interpretation and Surgical Pathology*, pp. 35–50, London: Springer London, 1981.

[34] M. Albanna and J. H. Holmes, *Skin tissue engineering and regenerative medicine.* Elsevier AP, 2016.

[35] J. Hicks, E. Wartchow, and G. Mierau, "Glycogen Storage Diseases: A Brief Review and Update on Clinical Features, Genetic Abnormalities, Pathologic Features, and Treatment," *Ultrastructural Pathology*, vol. 35, pp. 183–196, Oct. 2011.

[36] H. Mohan, A. Bal, and R. Aulakh, "Evaluation of skin biopsies for fungal infections: role of routine fungal staining," *Journal of Cutaneous Pathology*, vol. 35, pp. 1097–1099, Dec. 2008.

[37] J. F. A. McManus, "Histological and Histochemical Uses of Periodic Acid," *Stain Technology*, vol. 23, pp. 99–108, Jan. 1948.

[38] B. R. Masters, "Quantitative Phase Imaging of Cells and Tissues," *Journal of Biomedical Optics*, vol. 17, no. 2, p. 029901, 2012.

[39] P. Calzavara-Pinton, C. Longo, M. Venturini, R. Sala, and G. Pellacani, "Reflectance confocal microscopy for in vivo skin imaging†," *Photochemistry and Photobiology*, vol. 84, no. 6, pp. 1421–1430.

[40] J. Yao and L. V. Wang, "Photoacoustic microscopy," *Laser & Photonics Reviews*, vol. 7, no. 5, p. 758–778, 2013.

[41] F. H. KASTEN, "1 - the origins of modern fluorescence microscopy and fluorescent probes," in *Cell Structure and Function by Microspectrofluorometry* (E. KOHEN, ed.), pp. 3–50, Academic Press, 1989.

[42] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Science & Applications*, vol. 8, p. 23, Dec. 2019.

[43] S. Kretschmer, M. Pieper, G. Hüttmann, T. Bölke, B. Wollenberg, L. M. Marsh, H. Garn, and P. König, "Autofluorescence multiphoton microscopy for visualization of tissue morphology and cellular dynamics in murine and human airways," *Laboratory Investigation*, vol. 96, pp. 918–931, Aug. 2016.

[44] P. Kedarisetti, N. J. M. Haven, B. S. Restall, M. T. Martell, and R. J. Zemp, "Label-free lipid contrast imaging using non-contact near-infrared photoacoustic remote sensing microscopy," *Optics Letters*, vol. 45, p. 4559, Aug. 2020.

[45] N. J. M. Haven, K. L. Bell, P. Kedarisetti, J. D. Lewis, and R. J. Zemp, "Ultraviolet photoacoustic remote sensing microscopy," *Optics Letters*, vol. 44, p. 3586, July 2019.

[46] K. Bell, S. Abbasi, D. Dinakaran, M. Taher, G. Bigras, F. K. H. van Landeghem, J. R. Mackey, and P. Haji Reza, "Reflection-mode virtual histology using photoacoustic remote sensing microscopy," *Scientific Reports*, vol. 10, p. 19121, Dec. 2020.

[47] A. G. Bell, "The Production of Sound by Radiant Energy," *Science*, vol. 2, no. 48, pp. 242–253, 1881. Publisher: American Association for the Advancement of Science.

[48] L. V. Wang and S. Hu, "Photoacoustic Tomography: In Vivo Imaging from Organelles to Organs," *Science*, vol. 335, pp. 1458–1462, Mar. 2012.

[49] L. V. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," *Nature Methods*, vol. 13, pp. 627–638, Aug. 2016.

[50] B. Park, D. Oh, J. Kim, and C. Kim, "Functional photoacoustic imaging: From nano- and micro- to macro-scale," *Nano Convergence*, vol. 10, no. 1, 2023.

[51] C. J. Pavlin, M. D. Sherar, and F. S. Foster, "Subsurface Ultrasound Microscopic Imaging of the Intact Eye," *Ophthalmology*, vol. 97, pp. 244–250, Feb. 1990.

[52] R. H. Silverman, J. A. Ketterling, J. Mamou, and D. J. Coleman, "Improved High-Resolution Ultrasonic Imaging of the Eye," *Archives of Ophthalmology*, vol. 126, pp. 94–97, Jan. 2008.

[53] R. H. Silverman, "High-resolution ultrasound imaging of the eye – a review," *Clinical & Experimental Ophthalmology*, vol. 37, no. 1, pp. 54–67, 2009. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1442-9071.2008.01892.x.

[54] M. Li, Y. Tang, and J. Yao, "Photoacoustic tomography of blood oxygenation: A mini review," *Photoacoustics*, vol. 10, pp. 65–73, May 2018.

[55] W. Choi, E.-Y. Park, S. Jeon, and C. Kim, "Clinical photoacoustic imaging platforms," *Biomedical Engineering Letters*, vol. 8, no. 2, p. 139–155, 2018.

[56] S. Jeon, J. Kim, D. Lee, J. W. Baik, and C. Kim, "Review on practical photoacoustic microscopy," *Photoacoustics*, vol. 15, p. 100141, Sept. 2019.

[57] V. T. C. Tsang, X. Li, and T. T. W. Wong, "A Review of Endogenous and Exogenous Contrast Agents Used in Photoacoustic Tomography with Different Sensing Configurations," *Sensors*, vol. 20, p. 5595, Jan. 2020. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.

[58] D.-K. Yao, K. Maslov, K. K. Shung, Q. Zhou, and L. V. Wang, "In vivo label-free photoacoustic microscopy of cell nuclei by excitation of DNA and RNA," *Optics letters*, vol. 35, pp. 4139–4141, Dec. 2010.

[59] A. Sun, A. Sun, A. Sun, H. Guo, H. Guo, H. Guo, Q. Gan, L. Yang, Q. Liu, and L. Xi, "Evaluation of visible NIR-I and NIR-II light penetration for photoacoustic imaging in rat organs," *Optics Express*, vol. 28, pp. 9002–9013, Mar. 2020. Publisher: Optical Society of America.

[60] G. C. Langhout, D. J. Grootendorst, O. E. Nieweg, M. W. J. M. Wouters, J. A. van der Hage, J. Jose, H. van Boven, W. Steenbergen, S. Manohar, and T. J. M. Ruers, "Detection of Melanoma Metastases in Resected Human Lymph Nodes by Noninvasive Multispectral Photoacoustic Imaging," *International Journal of Biomedical Imaging*, vol. 2014, p. e163652, June 2014. Publisher: Hindawi.

[61] P. K. Upputuri and M. Pramanik, "Recent advances in photoacoustic contrast agents for in vivo imaging," *WIREs Nanomedicine and Nanobiotechnology*, vol. 12, no. 4, p. e1618, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wnan.1618.

[62] X. Wang, Y. Pang, G. Ku, X. Xie, G. Stoica, and L. V. Wang, "Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain," *Nature Biotechnology*, vol. 21, pp. 803–806, July 2003. Number: 7 Publisher: Nature Publishing Group.

[63] T. T. W. Wong, R. Zhang, P. Hai, C. Zhang, M. A. Pleitez, R. L. Aft, D. V. Novack, and L. V. Wang, "Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy," *Science Advances*, vol. 3, p. e1602168, May 2017. Publisher: American Association for the Advancement of Science Section: Research Article.

[64] M. Mehrmohammadi, S. J. Yoon, D. Yeager, and S. Y. Emelianov, "Photoacoustic Imaging for Cancer Detection and Staging," *Current molecular imaging*, vol. 2, pp. 89–105, Mar. 2013.

[65] W. Song, Q. Wei, L. Feng, V. Sarthy, S. Jiao, X. Liu, and H. F. Zhang, "Multimodal photoacoustic ophthalmoscopy in mouse," *Journal of Biophotonics*, vol. 6, no. 6-7, pp. 505–512, 2013. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jbio.201200061.

[66] X. Liu, T. Liu, R. Wen, Y. Li, C. A. Puliafito, H. F. Zhang, and S. Jiao, "Optical coherence photoacoustic microscopy for in vivo multimodal retinal imaging," *Optics Letters*, vol. 40, p. 1370, Apr. 2015.

[67] V. P. Nguyen, Y. Li, M. Aaberg, W. Zhang, X. Wang, and Y. M. Paulus, "In Vivo 3D Imaging of Retinal Neovascularization Using Multimodal Photoacoustic Microscopy and Optical Coherence Tomography Imaging," *Journal of Imaging*, vol. 4, p. 150, Dec. 2018. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

[68] S. Jeon, J. Kim, J. Yun, and C. Kim, "Non-destructive photoacoustic imaging of metal surface defects," *Journal of Optics*, vol. 18, p. 114001, Nov. 2016.

[69] G. J. Tserevelakis, I. Vrouvaki, P. Siozos, K. Melessanaki, K. Hatzigiannakis, C. Fotakis, and G. Zacharakis, "Photoacoustic imaging reveals hidden underdrawings in paintings," *Scientific Reports*, vol. 7, p. 747, Apr. 2017. Number: 1 Publisher: Nature Publishing Group.

[70] X. Wang, G. Ku, M. A. Wegiel, D. J. Bornhop, G. Stoica, and L. V. Wang, "Noninvasive photoacoustic angiography of animal brains in vivo with near-infrared light and an optical contrast agent," *Optics Letters*, vol. 29, p. 730, Apr. 2004.

[71] J. Yao, K. Maslov, S. Hu, and L. V. Wang, "Evans blue dye-enhanced capillary-resolution photoacoustic microscopy in vivo," *Journal of Biomedical Optics*, vol. 14, no. 5, 2009.

[72] L. V. Wang and H.-i. Wu, *Biomedical optics: principles and imaging*. Hoboken, N.J: Wiley-Interscience, 2007. OCLC: ocm71427127.

[73] Y. Zhou, J. Yao, and L. V. Wang, "Tutorial on photoacoustic tomography," *Journal of Biomedical Optics*, vol. 21, June 2016.

[74] L. V. Wang, "Tutorial on Photoacoustic Microscopy and Computed Tomography," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 14, pp. 171–179, Jan. 2008. Conference Name: IEEE Journal of Selected Topics in Quantum Electronics.

[75] S.-L. Chen, L. J. Guo, and X. Wang, "All-optical photoacoustic microscopy," *Photoacoustics*, vol. 3, pp. 143–150, Dec. 2015.

[76] J. Yao and L. V. Wang, "Photoacoustic Microscopy," *Laser & photonics reviews*, vol. 7, Sept. 2013.

[77] Z. Hosseinaee, J. A. Tummon Simmons, and P. H. Reza, "Dual-Modal Photoacoustic Imaging and Optical Coherence Tomography [Review]," *Frontiers in Physics*, vol. 8, 2021. Publisher: Frontiers.

[78] J. Xia, J. Yao, and L. V. Wang, "Photoacoustic tomography: principles and advances," *Electromagnetic waves (Cambridge, Mass.)*, vol. 147, pp. 1–22, 2014.

[79] C. Lee, J. Y. Kim, and C. Kim, "Recent Progress on Photoacoustic Imaging Enhanced with Microelectromechanical Systems (MEMS) Technologies," *Micromachines*, vol. 9, p. 584, Nov. 2018. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

[80] E. M. Strohm, M. J. Moore, and M. C. Kolios, "Single Cell Photoacoustic Microscopy: A Review," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, pp. 137–151, May 2016. Conference Name: IEEE Journal of Selected Topics in Quantum Electronics.

[81] E. Strohm, E. S. L. Berndl, and M. Kolios, "Probing Red Blood Cell Morphology Using High-Frequency Photoacoustics," *Biophysical Journal*, vol. 105, pp. 59–67, July 2013.

[82] A. Danielli, K. I. Maslov, A. Garcia-Uribe, A. M. Winkler, C. Li, L. Wang, Y. Chen, G. W. D. Ii, and L. V. Wang, "Label-free photoacoustic nanoscopy," *Journal of Biomedical Optics*, vol. 19, p. 086006, Aug. 2014. Publisher: International Society for Optics and Photonics.

[83] J. Staley, P. Grogan, A. K. Samadi, H. Cui, M. S. Cohen, and X. Yang, "Growth of melanoma brain tumors monitored by photoacoustic microscopy," *Journal of Biomedical Optics*, vol. 15, p. 040510, July 2010. Publisher: International Society for Optics and Photonics.

[84] P. Zhang, L. Li, L. Lin, P. Hu, J. Shi, Y. He, L. Zhu, Y. Zhou, and L. V. Wang, "High-resolution deep functional imaging of the whole mouse brain by photoacoustic computed tomography in vivo," *Journal of Biophotonics*, vol. 11, no. 1, p. e201700024, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jbio.201700024.

[85] M. R. Chatni, J. Xia, R. Sohn, K. Maslov, Z. Guo, Y. Zhang, K. Wang, Y. Xia, M. Anastasio, J. Arbeit, and L. V. Wang, "Tumor glucose metabolism imaged in vivo in small animals with whole-body photoacoustic computed tomography," *Journal of Biomedical Optics*, vol. 17, July 2012.

[86] J. Chan, Z. Zheng, K. Bell, M. Le, P. H. Reza, and J. T. Yeow, "Photoacoustic Imaging with Capacitive Micromachined Ultrasound Transducers: Principles and Developments," *Sensors (Basel, Switzerland)*, vol. 19, Aug. 2019.

[87] X. L. Deán-Ben, G. A. Pang, F. Montero de Espinosa, and D. Razansky, "Non-contact optoacoustic imaging with focused air-coupled transducers," *Applied Physics Letters*, vol. 107, p. 051105, Aug. 2015. Publisher: American Institute of Physics.

[88] Z. Hosseinaee, M. Le, K. Bell, and P. H. Reza, "Towards non-contact photoacoustic imaging [review]," *Photoacoustics*, vol. 20, p. 100207, Dec. 2020.

[89] J. Kim, D. Lee, U. Jung, and C. Kim, "Photoacoustic imaging platforms for multimodal imaging," *Ultrasonography*, vol. 34, pp. 88–97, Apr. 2015.

[90] S. Jiao, M. Jiang, J. Hu, A. Fawzi, Q. Zhou, K. K. Shung, C. A. Puliafito, and H. F. Zhang, "Photoacoustic ophthalmoscopy for in vivo retinal imaging," *Optics Express*, vol. 18, pp. 3967–3972, Feb. 2010.

[91] T. Berer, E. Leiss-Holzinger, A. Hochreiner, J. Bauer-Marschallinger, and A. Buchsbaum, "Multimodal noncontact photoacoustic and optical coherence tomography

imaging using wavelength-division multiplexing," *Journal of Biomedical Optics*, vol. 20, p. 46013, Apr. 2015.

[92] J. Yao, "When pressure meets light: detecting the photoacoustic effect at the origin," *Light: Science & Applications*, vol. 6, pp. e17062–e17062, June 2017. Number: 6 Publisher: Nature Publishing Group.

[93] K. L. Bell, P. H. Reza, and R. J. Zemp, "Real-time functional photoacoustic remote sensing microscopy," *Optics Letters*, vol. 44, pp. 3466–3469, July 2019. Publisher: Optical Society of America.

[94] N. Pellegrino, B. R. Ecclestone, D. Dinakaran, F. Van Landeghem, P. Fieguth, and P. Haji Reza, "Time-domain feature extraction for target specificity in photoacoustic remote sensing microscopy," *Optics Letters*, vol. 47, p. 3952, Aug. 2022.

[95] S. H. Jack, D. B. Hann, and C. A. Greated, "Influence of the acousto-optic effect on laser Doppler anemometry signals," *Review of Scientific Instruments*, vol. 69, pp. 4074–4081, Dec. 1998.

[96] P. H. Reza, K. Bell, W. Shi, J. Shapiro, and R. J. Zemp, "Deep non-contact photoacoustic initial pressure imaging," *Optica*, vol. 5, p. 814, July 2018.

[97] J. E. D. Tweel, B. R. Ecclestone, M. Boktor, J. A. T. Simmons, P. Fieguth, and P. H. Reza, "Virtual Histology with Photon Absorption Remote Sensing using a Cycle-Consistent Generative Adversarial Network with Weakly Registered Pairs," 2023. Publisher: arXiv Version Number: 1.

[98] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018.

[99] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2017.

[100] Y. Wu, X. Wang, Y. Li, H. Zhang, X. Zhao, and Y. Shan, "Towards vivid and diverse image colorization with generative color prior," 2022.

[101] K. Nazeri, E. Ng, and M. Ebrahimi, "Image colorization using generative adversarial networks," in *Articulated Motion and Deformable Objects*, pp. 85–94, Springer International Publishing, 2018.

[102] Y. Wang, M. Xia, L. Qi, J. Shao, and Y. Qiao, "Palgan: Image colorization with palette generative adversarial networks," 2022.

[103] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv:1411.1784 [cs, stat]*, Nov. 2014. arXiv: 1411.1784.

[104] R. Yan, Q. He, Y. Liu, P. Ye, L. Zhu, S. Shi, J. Gou, Y. He, T. Guan, and G. Zhou, "Unpaired virtual histological staining using prior-guided generative adversarial networks," *Computerized Medical Imaging and Graphics*, vol. 105, p. 102185, 2023.

[105] Y. Liang, D. Lee, Y. Li, and B.-S. Shin, "Unpaired medical image colorization using generative adversarial network," *Multimedia Tools and Applications*, Jan. 2021.

[106] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," 2016.

[107] J. Shlens, "A tutorial on principal component analysis," 2014.

[108] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *Journal of Educational Psychology*, vol. 24, no. 6, p. 417–441, 1933.

[109] R. Bro and A. K. Smilde, "Principal component analysis," *Anal. Methods*, vol. 6, pp. 2812–2831, 2014.

[110] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[111] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (L. M. L. Cam and J. Neyman, eds.), vol. 1, pp. 281–297, University of California Press, 1967.

[112] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[113] N. Pellegrino, P. W. Fieguth, and P. Haji Reza, "K-Means for noise-insensitive multidimensional feature learning," *Pattern Recognition Letters*, vol. 170, pp. 113–120, June 2023.

[114] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, 2014.

[115] M. Lotfollahi, D. Daeinejad, S. Berisha, and D. Mayerich, "Digital staining of high-resolution ftir spectroscopic images," *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018.

[116] D. Li, H. Hui, Y. Zhang, W. Tong, F. Tian, X. Yang, J. Liu, Y. Chen, and J. Tian, "Deep learning for virtual histological staining of bright-field microscopic images of unlabeled carotid artery tissue," *Molecular Imaging and Biology*, vol. 22, no. 5, p. 1301–1309, 2020.

[117] Z. Chen, W. Yu, I. H. M. Wong, and T. T. W. Wong, "Deep-learning-assisted microscopy with ultraviolet surface excitation for rapid slide-free histological imaging," *Biomedical Optics Express*, vol. 12, p. 5920, Sept. 2021.

[118] W. Xie, Y. Chen, Y. Wang, L. Wei, C. Yin, A. K. Glaser, M. E. Fauver, E. J. Seibel, S. M. Dintzis, J. C. Vaughan, N. P. Reder, and J. T. C. Liu, "Microscopy with ultraviolet surface excitation for wide-area pathology of breast surgical margins," *Journal of Biomedical Optics*, vol. 24, p. 1, Feb. 2019.

[119] F. T. Nguyen, A. M. Zysk, E. J. Chaney, J. G. Kotynek, U. J. Oliphant, F. J. Bellafiore, K. M. Rowland, P. A. Johnson, and S. A. Boppart, "Intraoperative Evaluation of Breast Tumor Margins with Optical Coherence Tomography," *Cancer Research*, vol. 69, pp. 8790–8796, Nov. 2009.

[120] J. L. Fine, L. Kagemann, G. Wollstein, H. Ishikawa, and J. S. Schuman, "Direct Scanning of Pathology Specimens Using Spectral Domain Optical Coherence Tomography: A Pilot Study," *Ophthalmic Surgery, Lasers and Imaging Retina*, vol. 41, Nov. 2010.

[121] W. Jung and S. A. Boppart, "Modern Trends in Imaging V: Optical Coherence Tomography for Rapid Tissue Screening and Directed Histological Sectioning," *Analytical Cellular Pathology*, vol. 35, no. 3, pp. 129–143, 2012.

[122] B. Bai, X. Yang, Y. Li, Y. Zhang, N. Pillar, and A. Ozcan, "Deep learning-enabled virtual histological staining of biological samples," *Light: Science & Applications*, vol. 12, no. 1, 2023.

[123] C. Zhang, Y. S. Zhang, D.-K. Yao, Y. Xia, and L. V. Wang, "Label-free photoacoustic microscopy of cytochromes," *Journal of Biomedical Optics*, vol. 18, p. 020504, Jan. 2013.

[124] T. T. W. Wong, R. Zhang, C. Zhang, H.-C. Hsu, K. I. Maslov, L. Wang, J. Shi, R. Chen, K. K. Shung, Q. Zhou, and L. V. Wang, "Label-free automated three-dimensional imaging of whole organs by microtomy-assisted photoacoustic microscopy," *Nature Communications*, vol. 8, p. 1386, Nov. 2017.

[125] A. K. Glaser, N. P. Reder, Y. Chen, E. F. McCarty, C. Yin, L. Wei, Y. Wang, L. D. True, and J. T. Liu, "Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens," *Nature Biomedical Engineering*, vol. 1, p. 0084, July 2017.

[126] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[127] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, p. 800, 2008.

[128] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, pp. 1398–1402 Vol.2, 2003.

[129] A. Obukhov and M. Krasnyanskiy, "Quality assessment method for gan based on modified metrics inception score and fréchet inception distance," in *Software Engineering Perspectives in Intelligent Systems*, pp. 102–114, 2020.

[130] 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)*, pp. 64–71, 2017.

[131] J. Salido, N. Vallez, L. González-López, O. Deniz, and G. Bueno, "Comparison of deep learning models for digital h&e staining from unpaired label-free multispectral microscopy images," *Computer Methods and Programs in Biomedicine*, vol. 235, p. 107528, 2023.

[132] Z. Xu, C. F. Moro, B. Bozóky, and Q. Zhang, "Gan-based virtual re-staining: A promising solution for whole slide image analysis," *CoRR*, vol. abs/1901.04059, 2019.

[133] G. Zhang, B. Ning, H. Hui, T. Yu, X. Yang, H. Zhang, J. Tian, and W. He, "Image-to-images translation for multiple virtual histological staining of unlabeled

human carotid atherosclerotic tissue," *Molecular Imaging and Biology*, vol. 24, no. 1, p. 31–41, 2021.

[134] G. Zhang, H. Hui, B. Ning, D. Dong, J. Tian, and W. He, "Self-attention based virtual staining for bright-field images of label-free human carotid atherosclerotic plaque tissue section," *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021.

[135] L. Shi, I. H. Wong, C. T. Lo, and T. T. Wong, "One-side virtual histological staining model for complex human samples," *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2022.

[136] R. Zhang, Y. Cao, Y. Li, Z. Liu, J. Wang, J. He, C. Zhang, X. Sui, P. Zhang, L. Cui, and et al., "Mvfstain: Multiple virtual functional stain histopathology images generation based on specific domain mapping," *Medical Image Analysis*, vol. 80, p. 102520, 2022.

[137] O. Cetin, M. Chen, P. Ziegler, P. Wild, and H. Koeppl, "Deep learning-based restaining of histopathological images," *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022.

[138] C. Fabelo, L. E. Selmic, P. Huang, J. P. Samuelson, J. K. Reagan, A. Kalamaras, V. Wavreille, G. L. Monroy, M. Marjanovic, and S. A. Boppart, "Evaluating optical coherence tomography for surgical margin assessment of canine mammary tumours," *Veterinary and Comparative Oncology*, vol. 19, pp. 697–706, Dec. 2021.

[139] R. Ha, L. C. Friedlander, H. Hibshoosh, C. Hendon, S. Feldman, S. Ahn, H. Schmidt, M. K. Akens, M. Fitzmaurice, B. C. Wilson, and V. L. Mango, "Optical coherence tomography: A novel imaging method for post-lumpectomy breast margin assessment—a multi-reader study," *Academic Radiology*, vol. 25, no. 3, pp. 279–287, 2018.

[140] B. J. Vakoc, D. Fukumura, R. K. Jain, and B. E. Bouma, "Cancer imaging by optical coherence tomography: preclinical progress and clinical potential," *Nature Reviews Cancer*, vol. 12, pp. 363–368, May 2012.

[141] Y. Huang, S. Wang, Q. Guo, S. Kessel, I. Rubinoff, L. L.-Y. Chan, P. Li, Y. Liu, J. Qiu, and C. Zhou, "Optical Coherence Tomography Detects Necrotic Regions and Volumetrically Quantifies Multicellular Tumor Spheroids," *Cancer Research*, vol. 77, pp. 6011–6020, Nov. 2017.

[142] A. Croce and G. Bottiroli, "Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis," *European Journal of Histochemistry*, Dec. 2014.

[143] K. de Haan, Y. Zhang, J. E. Zuckerman, T. Liu, Y. Rivenson, W. D. Wallace, and A. Ozcan, "Deep learning-based transformation of h&e stained tissue into special stains," *Conference on Lasers and Electro-Optics*, 2022.

[144] X. Yang, B. Bai, Y. Zhang, Y. Li, K. de Haan, T. Liu, and A. Ozcan, "Virtual stain transfer in histology via cascaded deep neural networks," *ACS Photonics*, vol. 9, no. 9, p. 3134–3143, 2022.

[145] Y.-C. Chen, X. Tan, Q. Sun, Q. Chen, W. Wang, and X. Fan, "Laser-emission imaging of nuclear biomarkers for high-contrast cancer screening and immunodiagnosis," *Nature Biomedical Engineering*, vol. 1, pp. 724–735, Sept. 2017.

[146] T. Imai, J. Shi, T. T. W. Wong, L. Li, L. Zhu, and L. V. Wang, "High-throughput ultraviolet photoacoustic microscopy with multifocal excitation," *Journal of Biomedical Optics*, vol. 23, p. 1, Mar. 2018.

[147] J. Shi, T. T. W. Wong, Y. He, L. Li, R. Zhang, C. S. Yung, J. Hwang, K. Maslov, and L. V. Wang, "High-resolution, high-contrast mid-infrared imaging of fresh biological samples with ultraviolet-localized photoacoustic microscopy," *Nature Photonics*, vol. 13, pp. 609–615, Sept. 2019.

[148] X. Li, L. Kang, C. T. Lo, V. T. Tsang, and T. T. Wong, "High-speed ultraviolet photoacoustic microscopy for histological imaging with virtual-staining assisted by deep learning," *Journal of Visualized Experiments*, no. 182, 2022.

[149] K. Bell, S. Abbasi, D. Dinakaran, M. Taher, G. Bigras, F. K. H. van Landeghem, J. R. Mackey, and P. Haji Reza, "Reflection-mode virtual histology using photoacoustic remote sensing microscopy," *Scientific Reports*, vol. 10, p. 19121, Dec. 2020.

[150] B. R. Ecclestone, Z. Hosseinaee, N. Abbasi, K. Bell, D. Dinakaran, J. R. Mackey, and P. Haji Reza, "Three-dimensional virtual histology in unprocessed resected tissues with photoacoustic remote sensing (PARS) microscopy and optical coherence tomography (OCT)," *Scientific Reports*, vol. 11, p. 13723, Dec. 2021.

[151] MathWorks, "Image processing toolbox," 2021.

[152] MathWorks, "Computer vision toolbox," 2021.

[153] A. Goshtasby, "Image registration by local approximation methods," *Image and Vision Computing*, vol. 6, pp. 255–261, Nov. 1988.

[154] J. C. Wolf and G. Maack, "Evaluating the credibility of histopathology data in environmental endocrine toxicity studies: Credibility of histopathology data in environmental studies," *Environmental Toxicology and Chemistry*, vol. 36, pp. 601–611, Mar. 2017.

[155] M. Zhang and Q. Ling, "Supervised pixel-wise gan for face super-resolution," *IEEE Transactions on Multimedia*, vol. 23, pp. 1938–1950, 2021.

[156] J. Salido, N. Vallez, L. González-López, O. Deniz, and G. Bueno, "Comparison of deep learning models for digital H&E staining from unpaired label-free multispectral microscopy images," *Computer Methods and Programs in Biomedicine*, vol. 235, p. 107528, 2023.

[157] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," 2018.

[158] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," 2020.

[159] M. García-Domínguez, C. Domínguez, J. Heras, E. Mata, and V. Pascual, "Neural style transfer and unpaired image-to-image translation to deal with the domain shift problem on spheroid segmentation," 2022.

[160] A. Alotaibi, "Deep generative adversarial networks for image-to-image translation: A review," *Symmetry*, vol. 12, no. 10, 2020.

[161] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2868–2876, 2017.

[162] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," 2018.

[163] Y. Li, S. Tang, R. Zhang, Y. Zhang, J. Li, and S. Yan, "Asymmetric GAN for unpaired image-to-image translation," *IEEE Transactions on Image Processing*, vol. 28, pp. 5881–5896, dec 2019.

[164] J. E. D. Tweel, B. R. Ecclestone, M. Boktor, D. Dinakaran, J. R. Mackey, and P. H. Reza, "Automated Whole Slide Imaging for Label-Free Histology using Photon Absorption Remote Sensing Microscopy," 2023. Publisher: arXiv Version Number: 2.

[165] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognition*, vol. 19, no. 6, pp. 459–466, 1986.

[166] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, p. e453, 2014.

[167] A. Ruifrok and D. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal Quant Cytol Histol*, vol. 23, Jan. 2001.

[168] A. Punnappurath, A. Abuolaim, A. Abdelhamed, A. Levinshtein, and M. S. Brown, "Day-to-night image synthesis for training nighttime neural isps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10769–10778, June 2022.

[169] A. Goshtasby, "Image registration by local approximation methods," *Image and Vision Computing*, vol. 6, pp. 255–261, Nov. 1988.

[170] P. Fieguth, *Statistical Image Processing and Multidimensional Modeling*. Information Science and Statistics, New York, NY: Springer New York, 2011.

[171] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[172] V. Jolly, M. Dwivedi, Y. Patel, T. Yadav, and S. Dhage, "Bringing Monochrome to Life: A GAN-based Approach to Colorizing Black and White Images," in *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, (Lonavla, India), pp. 1–6, IEEE, Apr. 2023.

[173] Y. Zhang, K. de Haan, Y. Rivenson, J. Li, A. Delis, and A. Ozcan, "Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue," *Light: Science & Applications*, vol. 9, no. 1, 2020.

[174] Z. Dadaci, F. Kılınç, T. T. Ozer, G. O. Sahin, N. O. Acir, and M. Borazan, "Periodic acid–schiff staining demonstrates fungi in chronic anterior blepharitis," *Eye*, vol. 29, no. 12, p. 1522–1527, 2015.

[175] Z. Yang and Z. Chen, "Learning from paired and unpaired data: Alternately trained cyclegan for near infrared image colorization," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 467–470, 2020.

[176] Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light: Science & Applications*, vol. 8, p. 23, Dec. 2019.

[177] M. Lapierre-Landry, J. Carroll, and M. C. Skala, "Imaging retinal melanin: a review of current technologies," *Journal of Biological Engineering*, vol. 12, p. 29, Dec. 2018.

[178] K. L. Bell, P. Hajireza, and R. J. Zemp, "Coherence-gated photoacoustic remote sensing microscopy," *Optics Express*, vol. 26, pp. 23689–23704, Sept. 2018. Publisher: Optical Society of America.

[179] X. Cai, L. Li, A. Krumholz, Z. Guo, T. N. Erpelding, C. Zhang, Y. Zhang, Y. Xia, and L. V. Wang, "Multi-Scale Molecular Photoacoustic Tomography of Gene Expression," *PLOS ONE*, vol. 7, p. e43999, Aug. 2012. Publisher: Public Library of Science.

[180] 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*, vol. 322, pp. 1857–1861, Dec. 2008.

[181] D. G. Eichberg, A. H. Shah, L. Di, A. M. Semonche, G. Jimsheleishvili, E. M. Luther, C. A. Sarkiss, A. D. Levi, S. H. Gultekin, R. J. Komotar, and M. E. Ivan, "Stimulated Raman histology for rapid and accurate intraoperative diagnosis of CNS tumors: prospective blinded study," *Journal of Neurosurgery*, vol. 134, pp. 137–143, Jan. 2021.

[182] 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*, vol. 1, p. 0027, Feb. 2017.

[183] E. Min, S. Ban, J. Lee, A. Vavilin, S. Baek, S. Jung, Y. Ahn, K. Park, S. Shin, S. Han, H. Cho, W. Lee-Kwon, J. Kim, C. J. Lee, and W. Jung, "Serial optical

coherence microscopy for label-free volumetric histopathology," *Scientific Reports*, vol. 10, p. 6711, Dec. 2020.

[184] S. Abbasi, M. Le, B. Sonier, D. Dinakaran, G. Bigras, K. Bell, J. R. Mackey, and P. Haji Reza, "All-optical Reflection-mode Microscopic Histology of Unstained Human Tissues," *Scientific Reports*, vol. 9, p. 13392, Dec. 2019.

[185] S. Soltani, A. Ojaghi, H. Qiao, N. Kaza, X. Li, Q. Dai, A. O. Osunkoya, and F. E. Robles, "Prostate cancer histopathology using label-free multispectral deep-uv microscopy quantifies phenotypes of tumor aggressiveness and enables multiple diagnostic virtual stains," *Scientific Reports*, vol. 12, no. 1, 2022.