

2023

Augmenting label-free imaging modalities with deep learning based digital staining

<https://hdl.handle.net/2144/46634>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
COLLEGE OF ENGINEERING

Dissertation

**AUGMENTING LABEL-FREE IMAGING MODALITIES
WITH DEEP LEARNING BASED DIGITAL STAINING**

by

SHIYI CHENG

B.E., University of Science and Technology of China, 2018
M.S., Boston University, 2023

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2023

© 2023 by
SHIYI CHENG
All rights reserved

Approved by

First Reader

Lei Tian, Ph.D.
Assistant Professor of Electrical and Computer Engineering
Assistant Professor of Biomedical Engineering

Second Reader

Eshed Ohn-Bar, Ph.D.
Assistant Professor of Electrical and Computer Engineering
Assistant Professor of Computer Science

Third Reader

David A. Boas, Ph.D.
Arthur G. B. Metcalf Chair
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering

Fourth Reader

Irving J. Bigio, Ph.D.
Professor of Biomedical Engineering
Professor of Electrical and Computer Engineering
Professor of Physics
Professor of Medicine

Fifth Reader

Ji-Xin Cheng, Ph.D.
Moustakas Chair Professor in Optoelectronics and Photonics
Professor of Electrical and Computer Engineering
Professor of Biomedical Engineering
Professor of Materials Science and Engineering
Professor of Chemistry
Professor of Physics

Acknowledgments

Financial support for this thesis is gratefully acknowledged from National Science Foundation (1846784), National Institute of Health (R01 NS108464, R01 EY032163).

I would like to express my sincere gratitude to my advisor, Professor Lei Tian, for his guidance and mentorship throughout my doctoral studies. His expertise and insights from diverse background and experiences have been pivotal in shaping my academic as well as personal growth. Ever since I joined his lab, he has been guiding me what to know and how to do research with unwavering patience. During my five years' Ph.D. journey, he has been providing me with exciting research opportunities and ideas to explore. His points of views have always been insightful and taught me more than just problem solving or research abilities. I learned a lot from his personality and energy. I am grateful for the opportunities and experiences provided by his lab and his continuous encouragement to push the boundaries of my research.

I'm also grateful to my committee members Professor Eshed Ohn-Bar, Professor David A. Boas, Professor Irving J. Bigio and Professor Ji-Xin Cheng for the valuable feedback on my Ph.D. work and thesis. The suggestions from Professor Ohn-Bar on the computer vision perspective are insightful and definitely triggering further line of research in the field. I would like to thank Professor Cheng for being supportive and showing great passion for my work and the fundamental problems I'm trying to solve. It has been a great honor and pleasure to be able to collaborate with Professor Boas and Professor Bigio on the cross-disciplinary project of digital staining OCT. Being among my committee as well as collaborators, their continuous support, understanding and insightful discussions have been extremely helpful for finding a path through so much ambiguity in a multi-disciplinary research collaboration.

I would also like to extend my gratitude to my collaborators who have helped me along the way. I enjoyed working with Professor Ji Yi for the first two years of

my Ph.D. on the collaboration of digital fluorescence labeling project. It's a great pleasure working with Professor Douglas L. Rosene, whose expertise on biology greatly strengthens the quality of my work. His heartfelt passion and patience came just at the right time when I didn't have a clue of what I'm about to do. Working with Professor Hui Wang and discussing interesting research problems have been an enlightening experience. Their collaboration, discussions, and shared expertise have greatly enriched my understanding and expanded the scope of my work. I would like to express my gratitude for my colleague collaborators Shuaibin Chang and Anna Novoseltseva for the support over the last two years. None of my work could be possible without their understanding and hard work. Their contributions and valuable insights have undoubtedly influenced the quality and depth of my research. I am honored to have had the privilege of working with them, and I am thankful for their dedication and commitment to advancing knowledge in our field.

I am grateful for being part of Lei's Computational Imaging System Lab and being able to work with a group of brilliant minds and warm hearts. I want to thank my senior members Dr. Alex Matlock, Dr. Waleed Tahir, Dr. Yujia Xue and Dr. Yunzhe Li for so many different perspectives I've gained from their diverse backgrounds. When I joined the lab, I knew nothing about the field and they helped me grow little by little. Among them, I want to give my sincere appreciation to Dr. Yujia Xue, who taught me everything from scratch - principles, algorithms, coding, PPT engineering, presenting and writing, literally everything. His warmest mentoring paves a way for me to smoothly transit from an undergrad to a mature researcher. He also offers me support and opportunities finding my career path and starting the next chapter of my life. It is a most amazing experience working with Dr. Yujia Xue and Dr. Yunzhe Li and participating their research. I appreciate the time wasted on discussing research ideas that we never get to try. I appreciate the happy lunch time

spent together on the 7th floor. I would also like to thank Joe Greene, Hao Wang, Jiabei Zhu, Qianwan Yang, Jeffrey Alido, Chang Liu, Guorong Hu, Dr. Ruipeng Guo and Dr. Shuying Li for the fun we had together. I can't wait to see their success doing the things they love.

I would like to dedicate my heartfelt appreciation and acknowledgment to the love of my life, Dr. Yunzhe Li, whose presence in my life has been nothing short of transformative. To Yunzhe, I extend my deepest thanks for your continuous company that have enriched my Ph.D. journey in immeasurable ways. You are my angel, my muse, my dream come true and my endless source of joy. I am lucky to have such a beautiful, charming, sweet, loving and interesting soul that cares for me. It may sound cliché but the day you walked into my life is like the light illuminated my world, then everything turned so beautiful. Before I met you, I'm a stubborn cynical old soul who wants to die young. I've never felt an actual flipping heart. I've never tasted the pure joy of caring for the one I love. I've never known staying together doing nothing could be just like heaven. I could never forget the days and nights we spent in Boston, along the Charles river, holding hands, watching snow, talking, walking, window shopping, laughing, crying, swearing, staying up, eating hotpot and arguing over dumplings. When I'm holding you in my arms, my heart melts and I feel so complete as if I'm holding my whole world, I wish time could just pass by till we both get old. Every moment spent with you is a moment I treasure. Your constant support during my depressed times has been a source of comfort and strength, when nothing but thinking of you can make me happy. I cannot imagine what will become of me without having you in my life. You truly changed my perspective on the world and made me learn how to live actually, how to love actually and how to become a better person. You have given my life a purpose and a direction that once appeared so blurry. Life is too short and changing the world is such a little thing. The dog

days are over, from now on, I only want to fill your life with my unconditional love. Let's ramble on and live the most boring life happily ever after.

Lastly, I am deeply indebted to my mother, Yuqin Shang, for her unwavering understanding, encouragement, and belief in my abilities. Her constant support and sacrifices have been instrumental in my pursuit of higher education. Growing up in a traditional small city, most of my childhood was spent at school where I was being constantly judged and pushed. Still, I could feel her protection, her genuine care and even indulgence to the best of her ability, in every possible way. She never ceases to encourage me to try, to practice and to excel in whatever field she believes I can do. She is a strong, tough woman who made it through the long way and basically raises me up all by herself. Although being traditional Chinese people, we never express our love for each other, but I love my mother from the bottom of my heart and I know she loves me in her own way. I can't wait to spend more time with her and care for her. I want to thank my father, Ruizhong Cheng, for developing my interests in math, science and rock n roll in my childhood. I wish you could be here to see how good I've done, see me graduate, celebrate and spend holidays together. I am truly fortunate to have such amazing families in my life, and their presence has played a pivotal role in shaping who I am today. I am forever grateful for their love, encouragement, and unwavering belief in my potential.

AUGMENTING LABEL-FREE IMAGING MODALITIES WITH DEEP LEARNING BASED DIGITAL STAINING

SHIYI CHENG

Boston University, College of Engineering, 2023

Major Professor: Lei Tian, Ph.D.

Assistant Professor of Electrical and Computer
Engineering

Assistant Professor of Biomedical Engineering

ABSTRACT

Label-free imaging modalities offer numerous advantages, such as the ability to avoid the time-consuming and potentially disruptive process of physical staining. However, one challenge that arises in label-free imaging is the limited ability to extract specific structural or molecular information from the acquired images. To overcome this limitation, a novel approach known as digital staining or digital labeling has emerged. Digital staining leverages the power of deep learning algorithms to virtually introduce labels or stains into label-free images, thereby enabling the extraction of detailed information that would typically require physical staining. The integration of digital staining with label-free imaging holds great promise in expanding the capabilities of imaging techniques, facilitating improved analysis, and advancing our understanding of biological systems at both the cellular and tissue level. In this thesis, I explore supervised and semi-supervised methodologies of digital staining and the applications in augmenting label-free imaging modalities, particularly in the context of cell imaging and brain imaging.

In the first part of the thesis, I demonstrate the novel integration of multi-contrast

dark-field reflectance microscopy and supervised deep learning to enable subcellular immunofluorescence labeling and cell cytometry from label-free imaging. By leveraging the rich structural information and sensitivity of reflectance microscopy, this method accurately predicts subcellular features without the need for physical staining. As a result of the use of a novel multi-contrast modality, the digital labeling approach demonstrates significant improvements over the state-of-the-art techniques, achieving up to $3\times$ prediction accuracy. In addition to fluorescence prediction, the method successfully reproduces single-cell level structural phenotypes related to cell cycles. The multiplexed readouts obtained through digital labeling enable accurate multi-parametric single-cell profiling across a large cell population.

In the second part, I investigated a novel digital staining optical coherence tomography (DS-OCT) modality combining advantages of serial sectioning OCT and semi-supervised deep learning and demonstrated several advantages for the application of 3D histological brain imaging. The DS model is trained using a semi-supervised learning framework that incorporates unpaired translation, a biophysical model, and cross-modality image registration, which manifests broad applicability to other weakly-paired bioimaging modalities. The DS model enables the translation of S-OCT images to Gallyas silver staining, providing consistent staining quality across different samples. I further show that DS enhances contrast across cortical layer boundaries and enables reliable cortical layer differentiation. Additionally, DS-OCT preserves 3D-geometry on centimeter-scale brain tissue blocks. My pilot study demonstrates promising results on other anatomical regions acquired from different S-OCT systems, highlighting its potential for generalization in various imaging contexts.

Overall, I investigate the problems of augmenting label-free imaging modalities with deep learning generated digital stains. I explored both supervised and semi-supervised methods for building novel DS frameworks. My work showcased two

important applications in the field of immunofluorescence cell imaging and 3D histological brain imaging. On the one hand, the integration of DS techniques with multi-contrast microscopy has the potential to enhance the throughput of single-cell imaging cytometry, and phenotyping. On the other hand, integrating DS techniques with S-OCT holds great potential for high-throughput human brain imaging, enabling comprehensive studies on the structure and function of the brain. Through the exploration, I aim to shed light on the impact of digital staining in the field of computational imaging and its implications for various scientific disciplines.

Contents

1	Introduction	1
2	Integrating Multi-contrast Reflectance Microscopy with Supervised Deep Learning Enables Label-free Immunofluorescence Prediction	8
2.1	Background	8
2.2	Principles and Overview	9
2.3	Method	11
2.3.1	Cell preparation and immunofluorescence staining	11
2.3.2	Oblique illumination-based reflectance microscopy platform	12
2.3.3	Image data acquisition	14
2.3.4	Data preprocessing	15
2.3.5	Neural network implementation	18
2.3.6	Quantitative evaluation of network prediction	19
2.3.7	Saliency map visualization	24
2.4	Results	25
2.4.1	Qualitative fluorescence prediction results	25
2.4.2	Quantitative fluorescence prediction results achieve state-of-the-art	30
2.5	Neural network saliency analysis	34
2.6	Summary	36
3	Digital Multiplexed Fluorescence Labeling for Single-cell Cytometry	38

3.1	Background	38
3.2	Overview	39
3.3	Methods	40
3.3.1	Digital cytometry analysis	40
3.3.2	Cell profile analysis	43
3.4	Results	45
3.4.1	Multiplexed prediction recovers biological accurate cellular structures	45
3.4.2	Cell profile analysis on multiplexed images allows phenotyping and quantitative cytometry	48
3.5	Summary	52
4	Integrating Semi-supervised Digital Staining with Serial Sectioning Optical Coherence Tomography	53
4.1	Background	53
4.2	Principles and Overview	54
4.3	Methods	57
4.3.1	Serial-sectioning OCT platform and data processing	57
4.3.2	Sample preparation and staining experiments	59
4.3.3	Image processing	60
4.3.4	Semi-supervised deep learning framework	64
4.3.5	Training details of all baseline models for comparisons	74
4.3.6	Customized quantitative metrics	76
4.3.7	Image analysis	78
4.4	Results	80
4.4.1	Digital staining by semi-supervised learning using weakly-paired images	80

4.4.2	Digital staining enhances mesoscopic brain structures and provides high staining uniformity	85
4.5	Summary	90
5	Pilot study of DS-OCT for the application of enhancing 3D histological brain imaging and neuropathology	91
5.1	Background	91
5.2	Overview	92
5.3	Image analysis	92
5.4	Results	93
5.4.1	Digital staining enables reliable cortical layer differentiation and layer thickness quantification	93
5.4.2	Volumetric digital staining on cubic centimeter-scale brain tissue	97
5.4.3	Generalization to unseen anatomical regions	99
5.5	Summary	103
6	Conclusions	105
6.1	Multiplexed digital fluorescence labeling for single-cell cytometry . . .	105
6.1.1	Summary	106
6.1.2	Discussion	106
6.1.3	Future work	108
6.2	Semi-supervised DS-OCT for 3D histological brain imaging	108
6.2.1	Summary	109
6.2.2	Discussion	109
6.2.3	Future work	111
6.3	General Discussions	111
	References	116
	Curriculum Vitae	127

List of Figures

2.1	Overview of the DL augmented label-free cytometry technique. (A) A multimodal LED-array reflectance microscope is developed to acquire co-registered label-free reflectance and fluorescence images. Reflectance images from oblique darkfield illumination and computed drDPC contain rich morphological information and are the multi-channel input to my DL model. Two-channel epi-fluorescence images are acquired on the same sample to serve as the ground truth for training the DL model. (B) Individual DL models are trained independently with paired label-free and IF images. The saliency map is used to reveal specific label-free features captured by the model to perform the transformation. (C) To perform digital multiplexed predictions, the same reflectance input is fed to each network and makes six different IF predictions in parallel.	10
2.2	drDPC imaging platform. Reflectance darkfield images are acquired using four different illumination patterns of the LED. Two extra channels of DPC are calculated and concatenated to the dark-field image stack as auxiliary information. Co-registered fluorescence images are acquired on the same setup.	13

2·3	Data preprocessing pipeline. (A) An example of Hoechst stained DNA image with the whole 1920×1080 -pixel FOV. (B) The image is processed by the flat-field cropping step. (C) An example of low-SNR raw measurement of the proliferation label that is contaminated by hot pixels and sensor noise. (D) The denoised proliferation-label image by the Noise2Void unsupervised deep learning algorithm. (E) An example of raw measurement of the endosome label that is primarily contaminated by hot pixels and with relatively low sensor noise. (F) The hot pixels are removed by morphological opening operation. (G) An example image and the histogram of measurement for the Golgi apparatus label that contains a constant background offset. (H) The processed Golgi label image and its histogram.	17
2·4	Neural network structure. Dense-blocks are added to facilitate efficient training. Skip connections are added to tunnel the high-frequency information from shallower layers to deeper layers with the same spatial scales. The lateral spatial size is marked by the number in the bottom of each block, and the layer depth size is marked by the number in the top of each block. The lateral sizes are compressed and decompressed by 2×2 maxpooling and upsampling layers, respectively.	19
2·5	Procedure of quantifying cell-level detection performance. Image processing pipeline for the quantification of four cell-level detection outcomes, including true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). All the detection performances are in terms of the concentration of target IF.	22

2·6	Saliency computation. The saliency map is determined by the gradient of the norm of the output with respect to the given input image. This shows an example of the calculated gradients of filter 10 in layer 10 and the averaged and smoothed gradient of all filters in layer 10 with respect to the input image I.	24
2·7	Visualization of the results from the six IF prediction networks. The rows show sample darkfield reflectance images from each input stack, the network’s IF prediction, and the ground-truth IF image, respectively. The columns show six IF labels covering nuclei (DNA), endosome, actin, Golgi apparatus, proliferation and apoptosis.	26
2·8	Additional individual prediction results. The rows show a sample reflectance image from each input stack, the network’s prediction, and the ground-truth IF image, respectively. The columns represent six IF labels including nuclei (DNA), endosome, actin, Golgi apparatus, proliferation and apoptosis.	27
2·9	Additional results. Rows: sample reflectance image, network prediction, ground-truth IF image. Columns: IF labels including nuclei (DNA), endosome, actin, Golgi apparatus, proliferation and apoptosis.	27
2·10	Additional individual prediction results.	28
2·11	Network prediction with and without drDPC input channels. The first two columns show examples of the darkfield reflectance and drDPC images used for DNA, endosome, actin, Golgi, proliferation and apoptosis prediction, respectively. The last three columns show the predicted labels without the drDPC input, with the drDPC input and the ground truth labels, respectively.	29

2.12	Quantitative evaluation of the DL prediction. (A) The image patch-wise PCCs of the predictions for nuclei (DNA), endosome, actin, and Golgi apparatus, which evaluates the pixel-level similarity between the regression-type predictions and ground-truth subcellular features. (B) The image patch-wise AUC of the proliferation and apoptosis predictions, which assess the pixel-level detection accuracy. (C) Quantitative evaluation of the cell-level detection performance of the proliferation and apoptosis predictions.	31
2.13	Quantification of network prediction with and without drDPC input channels. The prediction accuracy of DNA, endosome, Golgi degraded slightly with the additional drDPC input channels possibly because the drDPC morphological features do not add useful information for their predictions. The performance substantially improved for actin, proliferation and apoptosis with the additional drDPC input channels, indicating that the DPC morphological features contain rich information for their predictions.	33
2.14	Spatial variation quantification of prediction accuracy. (Top row) Red: the prediction of four ‘regression’ labels, including the DNA, endosome, actin, Golgi apparatus; Green: the ground truth overlay. (Bottom row) the corresponding PCC maps.	34

2.15	Saliency maps from each network across six cell batches with different staining conditions. The columns show the label-free input (the first darkfield reflectance channel) and the saliency maps for six different IF labels, including DNA (blue), endosome (red), actin (green), Golgi apparatus (yellow), proliferation (cyan), and apoptosis (magenta). The rows show the label-free input and the saliency maps from six cell batches under different staining conditions. The saliency maps show good consistency across different batches and highlight distinct morphological features.	35
3.1	Overview of the cell cytometry and profile analysis based on multiplexed digital fluorescence labeling. I evaluate several cellular features on the digitally multiplexed readouts using cell profiler. Several cellular metrics are extracted from the predicted IF labels and from the ground truth, and visualized in the violin plots.	39
3.2	Procedure of digital cytometry analysis. Image analysis pipeline for digital cytometry scatter plot. The pipeline applies to the DNA/proliferation image pairs from both the ground truth and the multiplexed predictions. The segmentation, ‘merge objects’ and ‘integrate intensity’ operations are implemented by the modules in CellProfiler.	41

3.3	Multiplexed prediction on six IF labels from the same label-free input. (A) Visualization of the Full-FOV multiplexed prediction including DNA (blue), endosome (red), actin (green), and Golgi apparatus (yellow), and (B) proliferation (cyan) and apoptosis (magenta) from the same reflectance input (grayscale). (C-R) Zoomed-in of DNA, Golgi apparatus, multiplexed, and proliferation predictions. White circles indicate representative cell morphology during different phases of the cell cycle, including (C-F) interphase, (G-J) metaphase, (K-N) anaphase, and (O-R) telophase.	46
3.4	Additional multiplexed prediction results from the endosome staining cell batch. (A) Visualization of the Full-FOV multiplexed prediction including DNA, endosome, actin, and Golgi apparatus. (B) Proliferation and apoptosis from the same reflectance input. The right-hand-side are zoomed-in of DNA, Golgi apparatus, multiplexed, and proliferation predictions. White circles indicate representative cell morphology during different phases of cell cycle.	47
3.5	Additional multiplexed prediction results from the proliferation staining cell batch.	48

3.6 Cell profile analysis on digital multiplexed IF staining. (A) Illustration of the cell cycle. (B-D) The scatter plots for the whole cell-level EdU (proliferating DNA) and Hoechst (DNA) concentrations from (B) the co-stained ground-truth, (C) the DL-prediction, and (D) the detection performance for the EdU predictions quantified by TP, TN, FP, and FN, across the entire cell population under the proliferation staining condition. (E-L) The comparisons of the statistics of eight single-cell profile metrics extracted from the entire cell population in the ground truth (GT) and the DL-predictions (pred), including (E) nuclear size, (F) DNA (nuclear) fluorescence intensity contrast, (G) cell (actin) size, (H) compactness of the actin, (I) NCR measured by the area ratio between the nuclei and actin, (J) endosome size, (K) eccentricity of the Golgi apparatus distribution, and (L) concentration of the Golgi apparatus. 49

3.7 Additional single-cell profile metrics. The comparisons of the statistics of twelve different single-cell profile metrics extracted from the entire cell population in the ground truth (GT) and DL-predictions (pred), including (A) nuclear size, (B) DNA (nuclear) fluorescence intensity variance, (C) nuclear perimeter, (D) cell (actin) size, (E) compactness of actin, (F) eccentricity of actin, (G) endosome size, (H) compactness of endosome, (I) eccentricity of endosome, (J) concentration of Golgi apparatus, (K) compactness of Golgi apparatus distribution, and (L) eccentricity of Golgi apparatus distribution. 51

4.1	Overview of the proposed OCT DS technique. (A) Data acquisition and DS model. S-OCT alternates between 3D imaging and tissue sectioning to acquire a stack of block-face OCT images, which are then processed to compute the scattering coefficient (OCT-SC) map stack. Sectioned sample slices are physically stained and imaged. The DS neural network is trained from a few weakly-aligned pairs of OCT-SC and Gallyas silver-stained images. (B) After the DS model is trained, it can perform inference on completely new slices of OCT-SC images for volumetric DS.	56
4.2	Image processing pipeline. (A) Preprocessing of PS images includes background removal, intensity normalization, resolution matching (down-sampling) and color transfer. (B) Preprocessing of OCT-SC images include edges & outliers masking, mask smoothing and maximum filtering. The outlier mask is determined by comparing pixel values to the local maximum shown on the right panel. (C) Post-processing of DS images. The second and third column shows results with sample masked and white matter (WM) region masked.	62

4.3 The training framework of my DS neural network model. (A) The backbone of the DS network G is built on the CUT framework, which combines contrastive learning and adversarial learning. The input is a 2D OCT-SC map X and the output is a digitally stained image $G(X)$ that is compared with a PS image Y from an adjacent slice. (B) Auxiliary pseudo-supervised learning task. The biophysical module computes the optical density $OD(Y)$ of the PS image Y , which is fed as an input to G . The digitally stained OD image $G(OD(Y))$ is compared with the original PS image Y during training. (C) Auxiliary unsupervised cross-modality image registration task. I alternate between optimizing G and a registration network R under different image scales. I fix R while updating G , which provides more informative supervision for R in the next iteration. I use patch-wise losses for training G , and WSI losses for training R 65

- 4.4 Detailed semi-supervised training framework of my DS model. (A) Semi-supervised training of DS generator combining contrastive learning, adversarial learning, pseudo-supervised learning and learnable registration. Patch-wise losses are formulated and combined in the total loss. Blue arrows and blue boxes indicate images are processed at the WSI scale. Black arrows, dotted black double-arrows and dotted black boxes indicate images are in patch-wise scale and are processed patch-wise. (B) Unsupervised cross-modality image registration. WSI losses are formulated between different image-modality pairs at two training stages. R is trained stand-alone at the pre-training stage, where similarity prior is used. At the fine-tuning stage, R is trained collaboratively with G using alternate optimization. Red arrows indicate the forwards pass during pre-training, and green arrows indicate the forward pass during fine-tuning. All images are processed as WSI scale in (B). 69
- 4.5 Model architecture details. The building blocks are convolutional blocks (ConvBlock) and residual blocks (ResBlock). The naming rules are as follows: Conv(i)S(j)-(k) represents an ii 2D convolution with k filters and stride j, where fractional stride size indicates transpose convolution; ConvBlock(i)S(j)-(k) represents a block consisting of an ii 2D convolution with k filters and stride j, followed by a batch normalization (BatchNorm) layer and a ReLu activation. ResBlock-(k) represents a ConvBlock3S1-k followed by a Conv3S1-k and BatchNorm; Linear-k is a fully connected layer with k hidden neurons. All ReLus in ConvBlocks of discriminator D are leaky ReLu. I use all leaky ReLus with a slope of 0.2. 73

4.6	Image analysis for fiber segmentation and quantification. (A) The mask of cortical layer IV/V/VI is segmented by combining an initial layer + white matter (WM) mask, a gray matter (GM) mask and brown color mask. The layer masks are directly used for quantitative evaluation. (B) Frangi filter and layer mask are combined to segment fiber tracts within layer IV/V/VI. Myelin fiber features such as lengths and diameters are extracted to compute histograms. Further quantitative metrics such as Jensen-Shannon divergence are computed from two histograms of extracted features.	79
4.7	Qualitative comparisons against previous baseline models. Results from four methods are visualized: CycleGAN, CUT, FastCUT, and Ours. PS images from adjacent cortex tissue sections are shown as references. Zoom-in regions are marked by the white rectangular boxes in WSIs. Scale bars are 1 mm.	82
4.8	Quantitative comparisons with baseline methods. (A) Box plots illustrate four pair-wise metrics, each represented by a unique color denoting the compared methods. The embedded numbers within the plots represent the median values for each metric. (B) Box plots present four unpaired metrics in five colors, with an additional gray color representing the metrics derived from PS images.	84
4.9	DS results on OCT-SC of tissue slices and comparisons with PS images. Cases include (A): ideal staining samples; (B): non-uniform staining and understanding samples. ROI 1, 3, 5, 7 are gyral crest regions and 2, 4, 6, 8 are sulcus regions. Scale bars are 1 mm.	86

4.10	Additional visualization of DS results. Cases include: Row 1: ideal staining; Row 2: under-staining sample; Row 3: over-staining sample; Row 4: non-uniform staining Scale bars are 1 mm.	89
5.1	Comparisons results of layer differentiation and thickness estimation in DS results. (A) The DS and PS WSIs from a cortex tissue section. (B) Zoom-in ROIs of inverted OCT-SC, DS and PS modalities marked in green and red boxes in (A) and normalized intensity profiles aggregates along white dotted lines. (C) Manually annotated layers IV/V/VI labeled in three colors and estimated local thickness. Statistics of thickness are visualized in box plot and grouped by gyral crest and sulcus regions. ROI is the zoom-in of dotted blue box from (A). .	94
5.2	Additional examples of improved layer differentiation in DS results. The DS and PS WSIs from four additional cortex tissue samples are shown on the left panel. Contrast enhanced grayscale images are shown in zoom-in ROIs. For these ROIs, The DS and PS ROIs are shown in order from left to right, while different ROIs are shown from top to bottom and marked by green and red boxes. Scale bars are 1 mm. . .	96
5.3	3D visualization and cross-sections views of the DS results on a large unseen tissue block. (A) The DS output images are stacked along the z-axis to render the whole digitally stained volume as well as segmented WM regions. (B) Orthogonal cross-sectional views of the DS volume. (C) Two zoom-in regions of vessel structures in yellow and green boxes from (A) are shown on the left. Three orthogonal maximum intensity projections (MIP) of the DS volume are shown on the right. Scale bars are 5 mm.	98

5.4 DS-OCT generalization performance on a hippocampus tissue slice. Examples of OCT-SC, DS and PS images (of adjacent sections) on one sample from Hippocampus region are shown. Each zoom-in ROIs correspond to numbered boxes in WSIs. Scale bars are 1 mm. 100

5.5 Additional examples of DS-OCT generalization performance. (A-D) tissue sections from four anatomical regions: cerebellum, somatosensory, BA21, and SupFrontal, respectively. In (A), red boxes are coarsely registered ROIs. Scale bars are 1 mm. 102

List of Abbreviations

AD	Alzheimer’s Disease
AUC	Area Under ROC Curve
BN	Batch Normalization
CC	Connected Component
CD	Color Difference
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Networks
CTE	Chronic Traumatic Encephalopathy
CUT	Contrastive Unpaired Translation
CycleGAN	Cycle-consistent Generative Adversarial Networks
DI	Deionized
DIC	Differential Interference Contrast
DL	Deep Learning
DoG	Difference-of-Gaussian
DPC	Differential Phase Contrast
drDPC	darkfield reflectance Differential Phase Contrast
DS	Digital Staining
FN	False Negative
FOV	Field of View
FP	False Positive
GAN	Generative Adversarial Networks
GM	Gray Matter
IF	Immunofluorescence
IOU	Intersection Over Union
JS	Jensen-Shannon divergence
MAD	Median Absolute Deviations
MF	Myelinated Fibers
MIP	Maximum Intensity Projection
MLP	Multi-layer Perceptron
MSE	Mean Squared Error
NA	Numerical Aperture
NCR	Nuclear-Cytoplasmic Ratio

NPCC	Negative Pearson Correlation Coefficient
OCT	Optical Coherence Tomography
OD	Optical Density
PCC	Pearson Correlation Coefficient
PS	Physical Staining
PSOCT	Polarization-sensitive Optical Coherence Tomography
QWP	Quarter Wave Plate
ReLU	Rectified Linear Unit
ResNet	Residual Neural Networks
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SC	Scattering Coefficients
SNR	Signal-to-Noise Ratio
S-OCT	Serial-sectioning Optical Coherence Tomography
TN	True Negative
TP	True Positive
TV	Total Variation
WM	White Matter
WSI	Whole Slide Image

Chapter 1

Introduction

Computational imaging is the process of recovering visual information from measurements using computation [1, 2, 3]. Over the years, the field of computational imaging has witnessed remarkable progress, driven by advancements in computing powers and computational techniques, such as the vast scale availability of high-performance computing platforms and the advances of deep learning algorithms [4, 5, 6, 7, 8, 9, 10]. Computational imaging has revolutionized the way we capture, process, and analyze visual information, enabling us to extract valuable insights from complex datasets. One particular area of interest within computational imaging is the development of label-free biomedical imaging techniques that eliminate the need for physical staining, offering advantages such as improved throughput and preservation of sample integrity [11, 12, 13, 14, 15, 16, 17, 18]. However, a challenge in label-free imaging has been the limited ability to extract specific structural or molecular information from acquired images. To address this limitation, deep-learning-based digital staining techniques have emerged as a promising solution. These techniques leverage the power of deep learning algorithms to virtually introduce labels or stains into label-free images, enabling the extraction of detailed information that would typically require physical staining [19]. By combining the strengths of label-free imaging and deep learning, digital staining approaches hold immense potential to expand the capabilities of label-free imaging techniques, enhance biomedical analysis, and advance our understanding of biological systems at both the cellular [20, 21, 22, 23] and tissue

level [24, 25, 26, 27]. In this thesis, I explore and further advance digital staining techniques, and discuss their implications for various biomedical imaging applications. By shedding light on the impact of digital staining, we aim to inspire further research and innovation in the field of computational imaging.

Label-free imaging techniques, such as differential phase contrast microscopy (DPC) [28, 29, 30] and optical coherence tomography (OCT) [31, 32, 33], provide valuable information about the morphology and optical properties of biological samples without the need for exogenous labels or dyes. However, these techniques have limitations that make digital staining necessary in some cases. On the one hand, label-free techniques primarily rely on intrinsic properties of the sample, such as refractive index [34, 30, 17] or scattering [11, 12, 13], to generate contrast. While this can reveal overall structural information, it often lacks the specificity to distinguish between different cellular components or molecular species. Digital staining, on the other hand, can selectively highlight specific targets of interest, such as DNA, organelles, or cortical layers by introducing artificial color or contrast enhancements. On the other hand, label-free techniques sometimes suffer from low contrast and sensitivity, especially when imaging samples with low refractive index variations or weak scattering properties that are disproportional to the stain concentration of certain structures. This can make it challenging to detect and analyze subtle structural features or low-abundance components. Digital staining can enhance contrast and sensitivity by applying algorithms or image processing techniques to amplify and highlight specific features, improving the visualization and analysis of the sample. Overall, label-free images can be difficult to interpret and often require domain expertise to annotate and analyze specific structures or components accurately. Digital staining can aid in the interpretation by providing familiar color-coded representations or contrast enhancements that facilitate the identification of different cellular

or anatomical structures. In my work, I further demonstrated that digital staining results can enable quantitative analysis by comparing the intensity of the stained regions using standard methods or by comparing the extracted cellular and histological features.

Digital staining (also known as virtual staining) is a computational technique that enhances and manipulates label-free images in order to reproduce the appearance of stained samples without the need for physical staining agents. As the most common approaches of digital staining methods, emerging deep learning algorithms have been successfully applied to provide dense labeling for label-free imaging modalities. The high expressivity of neural networks promises learning pixel-wise translation between different image domains. By training on a large dataset of stained and label-free images, deep learning models, such as convolutional neural networks (CNNs), learn to infer staining patterns based on the inherent correlations between them and can predict stained images from unlabeled images.

Existing deep learning based staining methods can be categorized into supervised approaches and unsupervised approaches. Supervised learning can be applied with carefully prepared co-registered image pairs, which has been demonstrated and validated on virtual fluorescence labeling and histological staining by previous work. The first use of a supervised CNN model demonstrated digital staining from live cell cultures with different immunofluorescence dyes in a shared optical path, by using phase microscopy and a U-Net model [21]. The use of fluorescence labeled antibodies for digital staining in live cells has seen extension into 3D digital staining [20] and detailed evaluation of virtual labeling of mitochondria in living cells [23]. Meanwhile, digital H&E staining and stain-to-stain transformation has been shown extensively and for a multitude of different organ samples based on phase, auto-fluorescence and pre-stained modalities [24, 25, 26, 35, 27, 19]. More recently, label-free OCT has

also been shown to successfully synthesize virtual H&E staining [36]. Besides the conceptual introduction of digital staining, the field was undoubtedly fueled by the introduction of powerful deep learning models for image labeling, such as U-Net [37] and residual neural networks (ResNet) [38, 39]. By leveraging the hierarchical nature of CNNs, these models can capture complex staining patterns and generate virtual stains that closely resemble the physical staining. U-Net, a variant of CNN architecture that has shown great success in various medical image analysis tasks [37], is particularly effective for tasks that involve segmentation or pixel-level labeling and staining. ResNet is a deep CNN architecture known for its ability to address the vanishing gradient problem and facilitate the training of very deep networks [38, 39], and has been employed to facilitate high-quality virtual stain generation. These architectures serve as backbones of the advancement of digital staining techniques using a basic supervised approach. However, most of the existing methods for digital fluorescence labeling rely on the brightfield transmittance modality which lacks sensitivity to high-resolution subcellular structures. Moreover, previous work is limited in the multiplicity of the digitally produced labels due to the spectral overlap in the fluorescence dye.

On the other hand, in the field of biomedical and neuroimaging, paired images are very hard or impossible to acquire, since the physical staining process generally introduces inevitable sample distortion and destruction. Considering the challenge of limited paired data, unsupervised learning strategy has been investigated to train totally-unpaired image-to-image translation models. By using the high-dimensional distribution matching ability of generative adversarial networks (GAN) [40, 41], ‘Pix2Pix’ model is first introduced in digital staining [42], while ‘stainGAN’ was concurrently introduced for stain normalization [43]. Cycle-consistent Generative Adversarial Networks (CycleGAN) [44] is a type of generative adversarial network (GAN) that has

been later applied to digital staining. Contrastive Unpaired Translation (CUT) [45] is an improvement over CycleGAN that uses contrastive learning to achieve better content preservation and superior training efficiency. These approaches have opened new applications that allowed for stain style transfer and digital staining without the need for pixel-aligned paired data [46, 47]. Without the constraint of acquiring pixel-aligned image data using specialized systems, digital staining techniques have been successfully extended on more diverse modalities such as confocal microscopy [48], ultraviolet photoacoustic microscopy [49] and cryosectioned whole-slide images [50]. However, the unsupervised models still lag behind supervised models in terms of accuracy due to lack of strong training constraints [19].

In this thesis, I present two digital staining techniques that addresses the above challenges in both supervised and unsupervised settings under the scenarios of digital fluorescence labeling and digital histological staining, respectively. I first showcase the novel use of multi-contrast dark field reflectance microscopy to improve single-cell immunofluorescence (IF) labeling and demonstrate the application of cell phenotyping and cytometry by multiplexed prediction. Then I explore a novel semi-supervised deep learning framework for digital histological staining under weakly-paired dataset and demonstrate the application of 3D histological brain imaging and image analysis.

In Chapter 2, I describe the innovative integration of multi-contrast dark-field reflectance microscopy and supervised deep learning to enables highly-accurate labeling of subcellular features. Leveraging the rich structural information and sensitivity offered by reflectance microscopy, the digital labeling approach showcases notable advancements compared to existing techniques, delivering prediction accuracy enhancement of up to three times.

In Chapter 3, I showcase the potential of multiplexed digital fluorescence labeling to facilitate cell cytometry directly from label-free imaging. While traditional

fluorescence-based imaging cytometry methods rely on fluorescence markers for identifying specific structures, they are limited in throughput due to the time-consuming labeling process. To address this challenge, a label-free technique has been developed that combines reflectance microscopy with deep learning-augmented digital labeling. Beyond single-modality fluorescence labeling, the digital multiplexed images correctly reproduce structural phenotypes at the single-cell level that are associated with cell cycles, such as Golgi twins, Golgi haze during mitosis, and DNA synthesis. Furthermore, the digital labeling technique provides multiplexed readouts, allowing for precise multi-parametric profiling of individual cells across a large cell population. My results show that my method can dramatically improve the throughput for imaging cytometry toward applications for phenotyping, pathology, and high-content screening.

In Chapter 4, I explore a new method called digital staining optical coherence tomography (DS-OCT) that combines serial sectioning OCT (S-OCT) and semi-supervised digital staining. The digital staining model I developed was trained using a semi-supervised learning approach, which incorporates contrastive unpaired translation, a biophysical model, and unsupervised cross-modality image registration. By employing the digital staining (DS) model on S-OCT images, the framework allows for the transformation of these images into Gallyas silver staining, ensuring consistent staining quality across diverse samples. The proposed method achieves superior performance over other baseline methods, including CycleGAN, CUT and FastCUT in terms of DS quality and accuracy.

In Chapter 5, I demonstrate the pilot study of DS-OCT for the application of 3D histological human brain imaging and image analysis. In the context of neuroscience, a major challenge lies in visualizing the structure of the human brain at different scales. Traditional histology techniques, while capable of revealing micro- and meso-

scale features, suffer from staining variability, tissue damage, and distortion that limit the accuracy of 3D reconstruction. My novel 3D imaging framework addresses this challenge by combining the advantages of S-OCT and digital staining. My study demonstrates that DS greatly improves contrast at the boundaries between cortical layers and enables accurate discrimination of cortical layers IV/V/VI. Furthermore, the framework allows for geometry-preserving 3D DS on brain tissue blocks measuring cubic centimeters, facilitating the visualization of mesoscopic vessel networks within the white matter. Notably, in the pilot study involving different S-OCT systems and anatomical regions, the framework showcases promising outcomes, suggesting its potential for broad utilization and adaptability across diverse imaging scenarios.

To briefly summarize, in this thesis, I focus on the problems of enhancing label-free imaging techniques by incorporating digital stains generated through deep learning. My contributions are on the investigation of both supervised and semi-supervised methods to develop innovative frameworks for digital staining. My work also contributes to demonstrate the application of these frameworks in two important areas: immunofluorescence cell imaging and 3D histological brain imaging. By integrating multi-contrast microscopy and digital staining, one can potentially improve the efficiency of single-cell imaging cytometry and phenotyping. Furthermore, the integration of digital staining with S-OCT shows promise for high-throughput histological brain imaging, enabling comprehensive studies on human brains.

In Chapter 6, I summarize and discuss several limitations, future work and potential impact of the two major parts of my work. In light of the general digital staining techniques, I further include general discussions on the limitations, bottlenecks and pitfalls as well as suggestions of potential solutions accordingly. I conclude my thesis by sharing a perspective of the future of digital staining techniques.

Chapter 2

Integrating Multi-contrast Reflectance Microscopy with Supervised Deep Learning Enables Label-free Immunofluorescence Prediction

2.1 Background

In this section, I aim to develop deep learning (DL) methods for augmenting label-free microscopy with dense digital fluorescence labels. Previous work has demonstrated digital fluorescence labeling using supervised learning with carefully prepared image pairs. The fluorescence labels can be generally acquired on the same setup that captures label-free images, which produces co-registered pairs. The first use of a supervised CNN model demonstrated digital cell staining with different immunofluorescence (IF) dyes in a shared optical path, by using phase microscopy and a U-Net model [21], which is then extended into 3D digital staining [20]. However, previous work primarily utilized transmittance bright field microscopic images as label-free modality, with relatively low sensitivity and specificity to fine subcellular structures. Therefore, augmenting the label-free imaging modality with multiple contrasts is important to improve the accuracy of digital staining models. Moreover, most previous work is limited in the multiplicity of the digitally produced labels due to the spectral overlap in the fluorescence dye or histology stain. The inherent bottleneck in the dataset has prevented the applicability of multiplexed virtual staining and limited

the imaging content throughput.

2.2 Principles and Overview

Our work relies on the premise that label-free scattering-based microscopy captures rich structural information and can be effective to characterize cell morphological features [51]. Brightfield, phase contrast and differential interference contrast (DIC) microscopy have been routinely used for observing and quantifying cell morphology [52]. Scattering-based microscopy and tomography techniques have been increasingly utilized to reconstruct cellular structures [53]. Of particular interest is the reflectance-mode microscopy that provides exquisite sensitivity in detecting nanoscale structural changes beyond the diffraction limit [54, 55, 56]. By capturing backscattering signals, reflectance imaging provides access to the highest spatial-frequency components in the reciprocal Fourier space and thus can provide higher structural contrast than the transmission techniques [51]. Indeed, our recent work shows that backscattering signals allow resolving finer details than the transmission counterparts [57, 58]. In this work, I further leverage the higher sensitivity provided by the reflectance-mode microscopy and demonstrate how enriched label-free information allows predicting highly accurate subcellular structural features.

The framework of this study is summarized in Fig. 2.1. The angle-dependent backscattering features are captured with darkfield oblique illumination and paired with IF images (Fig. 2.1A). Using the IF images as the ground truths, multiple DL models are independently trained for individual IF labels (Fig. 2.1B). Once all the models are trained, I perform digital multiplexing by feeding the same label-free input to each network and make different IF predictions in parallel (Fig. 2.1C). By doing so, multiple subcellular structures and cell states can be revealed simultaneously without physical labeling. While previous works have shown that DL models can disentangle

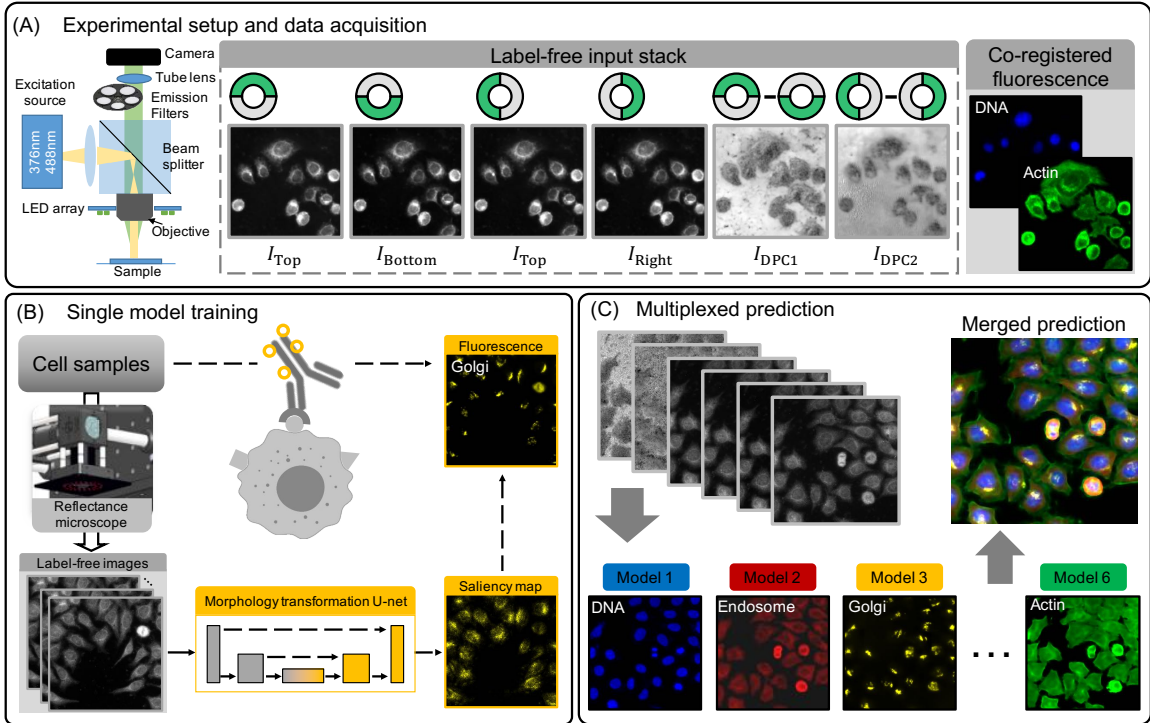


Figure 2.1: Overview of the DL augmented label-free cytometry technique. (A) A multimodal LED-array reflectance microscope is developed to acquire co-registered label-free reflectance and fluorescence images. Reflectance images from oblique dark-field illumination and computed drDPC contain rich morphological information and are the multi-channel input to my DL model. Two-channel epi-fluorescence images are acquired on the same sample to serve as the ground truth for training the DL model. (B) Individual DL models are trained independently with paired label-free and IF images. The saliency map is used to reveal specific label-free features captured by the model to perform the transformation. (C) To perform digital multiplexed predictions, the same reflectance input is fed to each network and makes six different IF predictions in parallel.

the complex structures captured in the label-free data and make in-silico fluorescence labeling with high accuracy [20, 21, 59] or holistically capture “hidden” structural features that are not easily perceived or described [60, 61, 62, 63, 64, 65, 66, 67, 68, 69], these results are fundamentally limited by the weak structural contrast from the transmission modes that contain only forward scattering information. By exploiting the enhanced resolution and sensitivity in the backscattering data, I demonstrate a dramatic increase in the fluorescence prediction accuracy with up to $3\times$ improvement

as compared to the current state-of-the-art.

A common criticism of DL-based methods is the “black-box” nature of these models [70]. To overcome this issue, I adapt the attention mechanism [71] to elucidate on the working mechanism of the DL model. I construct the saliency map that highlights the most important subcellular features contributing to each IF prediction by the network. My results show that the structural components in label-free reflectance input that correspond to the fluorescence labels can be correctly identified by the saliency map, and the “attention” is consistent across different cell batches when predicting all IF labels. This indicates that the network learns to extract the salient and specific structural information from the reflectance images matching the underlying subcellular components. In addition, the improved prediction accuracy is attributed to the enhanced resolution and sensitivity to subcellular structures from the backscattering information.

2.3 Method

2.3.1 Cell preparation and immunofluorescence staining

HeLa cells were cultured in a Dulbecco’s Modified Eagle Medium (DMEM) with 10% fetal bovine serum (Gibco, 10564011) and 5% penicillin streptomycin (Gibco, 15140122). The cells were trypsinized and passaged twice a week. Two days before the staining and imaging, cells were cultured on glass-bottom petri dishes (FluoroDish FD35-100) which were first treated with 10% poly-L-lysine (SigmaAldrich, RNBG0769) with PBS for ten mins in an incubator. The staining and imaging were performed on the glass-bottom dishes. I follow the standard IF staining protocols. In total, six IF stains are used to label DNA (Hoechst), actin (Phalloidin, Alexa Fluor 488 Phalloidin, Invitrogen, A12379), endosome (EEA1, Santa Cruz, sc-137130 AFF488), Golgi apparatus (GM130, Cell Signaling, 12480), proliferation

(EdU, Click-iT Plus EdU Alexa Fluor 488 kit, Invitrogen), and apoptosis (TUNEL, Click-iT TUNEL Alexa Fluor 488 kit, Invitrogen). The HeLa cells were first fixed with ice-cold methanol, washed three times (10 min each) in 0.05% PBST (0.05% Triton X-100 PBS solution), and incubated for 20 mins at room temperature in a blocking solution containing 0.25% Triton X-100 and 10% bovine serum albumin in PBS. Alexa-488 conjugated antibodies were diluted in the blocking solution with the recommended concentration by the manufacturers and incubated with cells to label the specific subcellular components (EEA1 for endosome, Phalloidin for actin). For actin staining, cells were fixed with ice-cold acetone to preserve the structures. For Golgi staining, a secondary antibody Anti-rabbit IgG (Santa Cruz, 4412S) was diluted in blocking solution and used to culture the cells for 1.5 hours at room temperature in dark. To stain cell proliferation and apoptosis, I used EdU and TUNEL assays, respectively, according to the recommended protocol by Invitrogen. The apoptosis was induced by culturing the cells with 1 μ M Staurosporine for 24 hours. In all the above stains, cell nuclei were counterstained with 1 \times Hoechst 33342.

2.3.2 Oblique illumination-based reflectance microscopy platform

The imaging platform is based on our recently developed LED-array reflectance microscope for capturing co-registered label-free reflectance and fluorescence images [57].

By flexibly controlling the LED patterns, this new platform enables capturing multiple angle-dependent backscattering contrasts in the darkfield without any mechanical switching. Based on our prior work [57], we heuristically optimize the illumination strategy and implement half-annulus LED patterns along four different orientations (including top, bottom, left, and right) (see Fig. 2.2). In addition, I compute the darkfield reflectance differential phase contrast (drDPC) based on the raw measurements. The raw oblique-illumination darkfield and drDPC images contain complementary structural contrasts. In particular, subcellular structures are shown

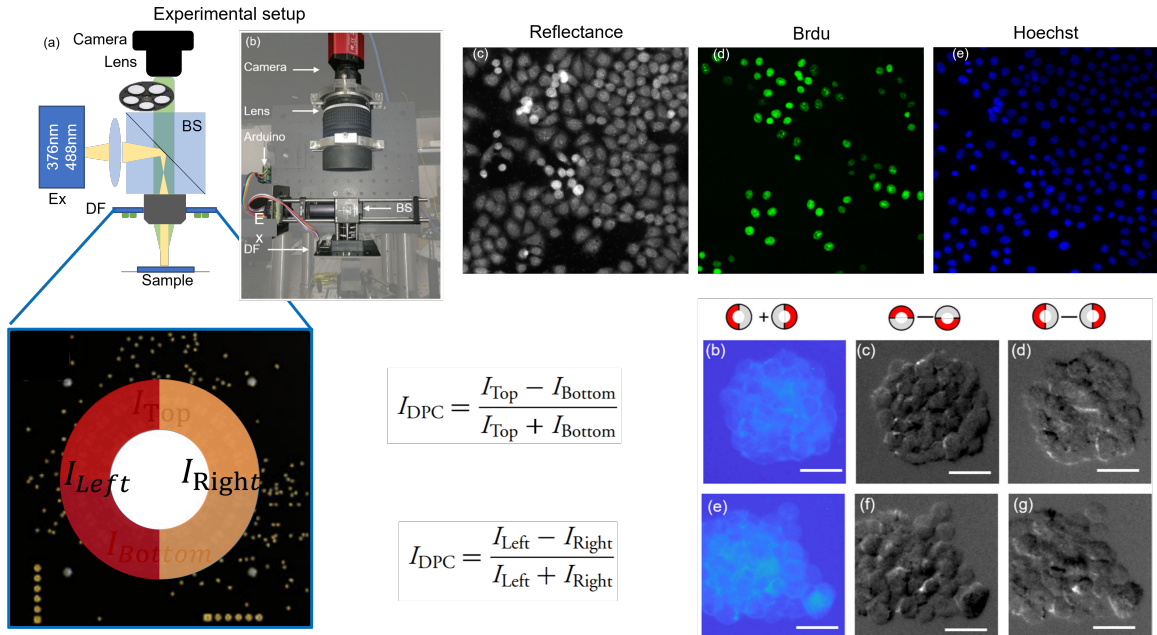


Figure 2-2: drDPC imaging platform. Reflectance darkfield images are acquired using four different illumination patterns of the LED. Two extra channels of DPC are calculated and concatenated to the dark-field image stack as auxiliary information. Co-registered fluorescence images are acquired on the same setup.

with high contrast in the raw darkfield measurements, including the nuclei, nucleoli, and hyper-reflective structures at the nuclear periphery. Cell membranes with sharp boundaries are highlighted in the drDPC images, with cytoplasm spreading on the substrate with thicker nuclei at the cells' centers. The extra cell topography information exhibited in the drDPC images are found particularly useful for predicting IF labels that are sensitive to morphology of the cell boundary. These label-free images are used as the multi-channel input to my DL model. On the same platform, two-channel epi-fluorescence images are concurrently acquired on the same sample to serve as the ground-truth for training my DL models (Fig. 2-1A). The significance of this new microscopy platform is that I capture enriched label-free information by multiple contrasts in the reflectance-mode. This empowers my label-free high-content cytometry technique to uncover highly sensitive and specific structural phenotypes at

the single-cell level across large cell populations.

2.3.3 Image data acquisition

We collect the data using our custom-built multimodal reflectance microscope [57], as shown in Fig. 2·1A. A custom-built LED array consisting of two LED rings is used for providing controllable darkfield illumination in reflection. We use commercially available LEDs (APTF1616SEEZGQBDC, Kingbright) that can provide three independent RGB color channels (central wavelength is 460, 515, and 630 nm, respectively). All the LEDs are individually addressable using two cascaded LED drivers (TLC5955, Texas Instruments). A microcontroller (Teensy 3.2, PJRC) provides the camera trigger signal through digital Input/Output pins and simultaneously controls the LED illumination pattern. The LED array is mounted around the objective lens ($10\times$ 0.3 NA, UPlanFL N, Olympus, Japan) using a 3D-printed adapter. The tube lens is a commercial SLR lens (Nikon AF DC-NIKKOR 135 mm f/2D) to maximize the FOV. The microscope provides an overall $7.5\times$ magnification. An sCMOS camera (CS2100M-USB, Thorlabs, 1920×1080 pixels, $5.04\ \mu\text{m}$ pixel size, 16-bit depth) is used to acquire the images. We capture four darkfield reflectance images by using half-annulus green LED patterns along different orientations, including top (I_{Top}), bottom (I_{Bottom}), left (I_{Left}), and right (I_{Right}). The exposure time is 700 ms. Two drDPC images along two orthogonal orientations are generated by

$$I_{\text{DPC}_1} = \frac{I_{\text{Top}} - I_{\text{Bottom}}}{I_{\text{Top}} + I_{\text{Bottom}}} \tag{2.1}$$

$$I_{\text{DPC}_2} = \frac{I_{\text{Left}} - I_{\text{Right}}}{I_{\text{Left}} + I_{\text{Right}}}$$

The fluorescence excitations are provided by two LED sources (M365LP1 and M470L4, Thorlabs, central wavelength 365 nm, 470 nm, respectively) combined with a dichroic mirror (DMLP425R, Thorlabs). The epi-fluorescence illumination is in-

roduced by a 50/50 beam splitter (CCM1-BS013, Thorlabs). The emission filters (MF460-60, MF525-39, Thorlabs) are placed on a filter wheel (CFW6, Thorlabs) for blue and green fluorescence emissions. Two-channel fluorescence images are acquired sequentially after acquiring the reflectance images. The exposure time is 400 ms for IF imaging. Specifically, the first green channel is for one of the five IF antibodies conjugated with the green fluorophores (Alexa 488) for endosome, actin, Golgi apparatus, proliferation, and apoptosis; the second blue channel is for the co-stained DNA. We capture 30 image stacks for each sample batch / IF stain.

2.3.4 Data preprocessing

The raw reflectance and fluorescence images are preprocessed before feeding into my deep neural networks for training. The preprocessing procedure consists of four steps, including flat-field cropping, image denoising, background correction, and intensity normalization.

Since the fluorescence excitation illumination is not evenly distributed across the entire rectangular FOV, I first perform flat-field cropping by using only the central 1000×1080 -pixel region for training, where the excitation is approximately uniform.

Second, I perform image denoising on the measurements. I apply two denoising approaches. In the first approach, I apply an unsupervised DL-based denoising algorithm, Noise2Void [72], to suppress the sensor noise present in the images. To do so, each 1000×1080 -pixel image is cropped into 256×256 -pixel patches. Each image patch is fed to a blind-spot network to perform denoising. After denoising, the patches are then stitched back together by alpha blending. This unsupervised denoising algorithm is found to be effective in removing unstructured, signal-independent noise, including the sensor noise and isolated hot pixels, in particular for measurements with low signal-to-noise ratios (SNRs). The whole training and inference (denoising) procedure takes 10 hours for processing the entire dataset containing 30 images.

I find this denoising procedure is only necessary for processing the Golgi and proliferation fluorescence images, as well as for the reflectance images for the actin prediction where the sensor noise severely corrupts the images. In the second approach, when the SNRs are sufficiently high for the measurements on other cell batches, I use a computationally more efficient morphology opening operation to remove the hot pixels in the fluorescence images under the assumption that hot pixels are isolated pixels with extreme intensity values. The opening operation takes a square kernel of size 2×2 pixels. This hot-pixel removal procedure takes 15 min to process the entire dataset containing 30 images.

Third, I perform background correction on the fluorescence images by eliminating the potential background bias across the batches. To do so, I calculate the histogram of each fluorescence image and denote the mode value (i.e. the most frequent value) as the constant background. This background of each fluorescence image is subtracted; the negative values from the subtraction are clipped to zero.

Fourth, I perform intensity normalization by normalizing the pixel values of both the input and output images to be between 0 and 1.

The data preprocessing steps are shown in Fig. 2-3. As an example in Fig. 2-3A, the image suffers from non-uniform fluorescence expression due to non-uniform excitation illumination, and is cropped to maintain only the central 1000×1080 -pixel region that retains approximately a uniform illumination in Fig. 2-3B. In Fig. 2-3D, both the sensor noise and the hot pixels are removed while maintaining the fluorescence signals with high fidelity, while in Fig. 2-3F the dominant hot pixels are effectively removed by a simple morphological operation. In Fig. 2-3H, the image is corrected by subtracting the background offset and normalizing the intensity range.

network takes 256×256 -pixel input images. Accordingly, I split the 30 pairs of images into 256×256 -pixel patches to generate the training and testing data. For

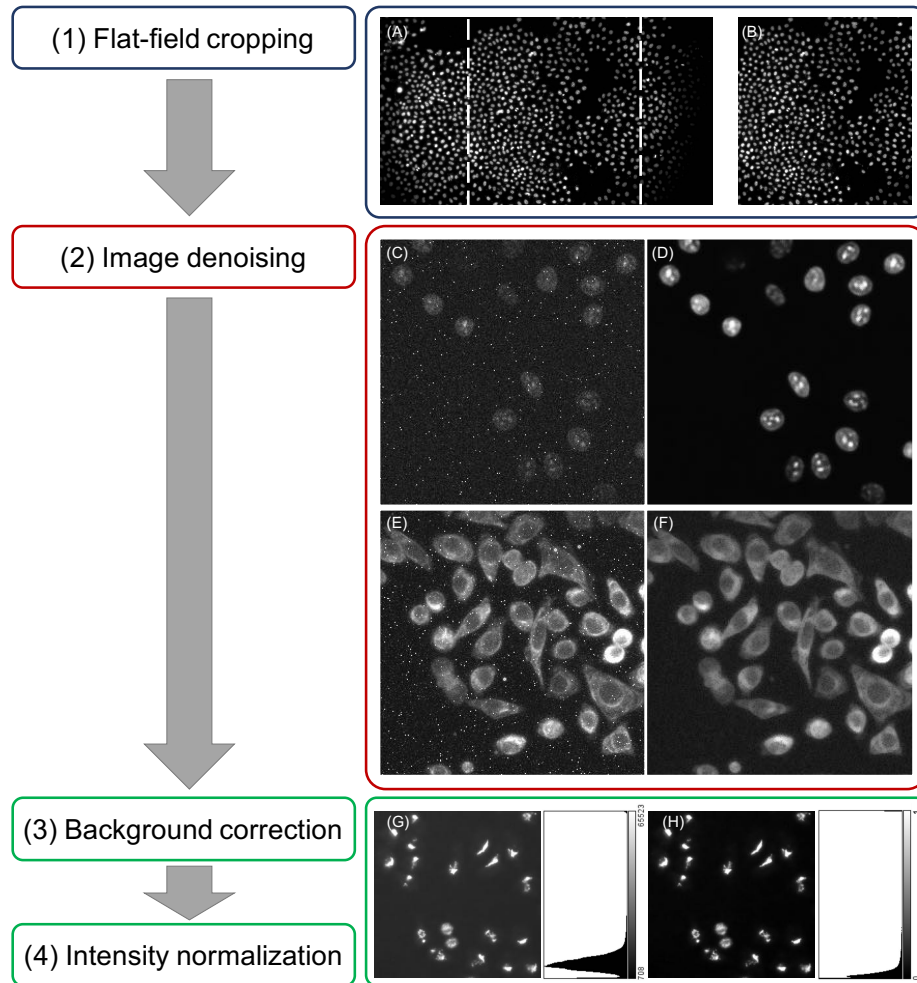


Figure 2-3: Data preprocessing pipeline. (A) An example of Hoechst stained DNA image with the whole 1920×1080 -pixel FOV. (B) The image is processed by the flat-field cropping step. (C) An example of low-SNR raw measurement of the proliferation label that is contaminated by hot pixels and sensor noise. (D) The denoised proliferation-label image by the Noise2Void unsupervised deep learning algorithm. (E) An example of raw measurement of the endosome label that is primarily contaminated by hot pixels and with relatively low sensor noise. (F) The hot pixels are removed by morphological opening operation. (G) An example image and the histogram of measurement for the Golgi apparatus label that contains a constant background offset. (H) The processed Golgi label image and its histogram.

each IF prediction, 512 training samples and 128 testing samples are randomly generated from the full field of view (FOV) image pairs. Each input stack consists of four different channels combining darkfield reflectance images from different oblique

illumination patterns (Fig. 2.1A). In addition, I construct the two-direction drDPC images from the darkfield images (Fig. 2.1A), which were found particularly effective for predicting the actin, proliferation, and apoptosis labels. When making the full-FOV predictions, I use the entire 1080×1920 -pixel reflectance images since they do not suffer from the non-uniform illumination issue.

2.3.5 Neural network implementation

I develop a convolutional neural network (CNN) to learn the highly complex non-linear mapping between the morphology information contained in the multi-channel reflectance images and the fluorescence labels. The overall structure of the neural network follows the U-Net architecture modified with the dense-block modules. The input takes 256×256 -pixel label-free images consisting of four darkfield reflectance images. For predicting the actin fluorescence labels, the input is expanded to six channels that include additionally two-directional drDPC images. For all the conditions, the output predicts a single-channel fluorescence image with 256×256 pixels.

The network is detailed in Fig. 2.4. It is primarily based on the encoder-decoder U-net architecture [37] and further incorporates the dense-blocks and skip-connections to enable high-resolution information prediction [10]. Starting with the multi-channel high-resolution label-free reflectance images, the encoder path gradually condenses the lateral spatial information into high-level feature maps with growing depths, and the decoder path reverses the process by recombining the information into feature maps with gradually increased lateral details. The information in adjacent feature maps transfers by convolving with 3×3 convolutional filters. Each dense-block contains multiple layers, in which each layer consists of batch normalization (BN), the rectified linear unit (ReLU) nonlinear activation, and convolution (conv) with 24 filters. I also added drop out layers with 0.5 dropout rate after the central two dense-blocks to mitigate overfitting. The input of the preprocessed 256×256 -pixel

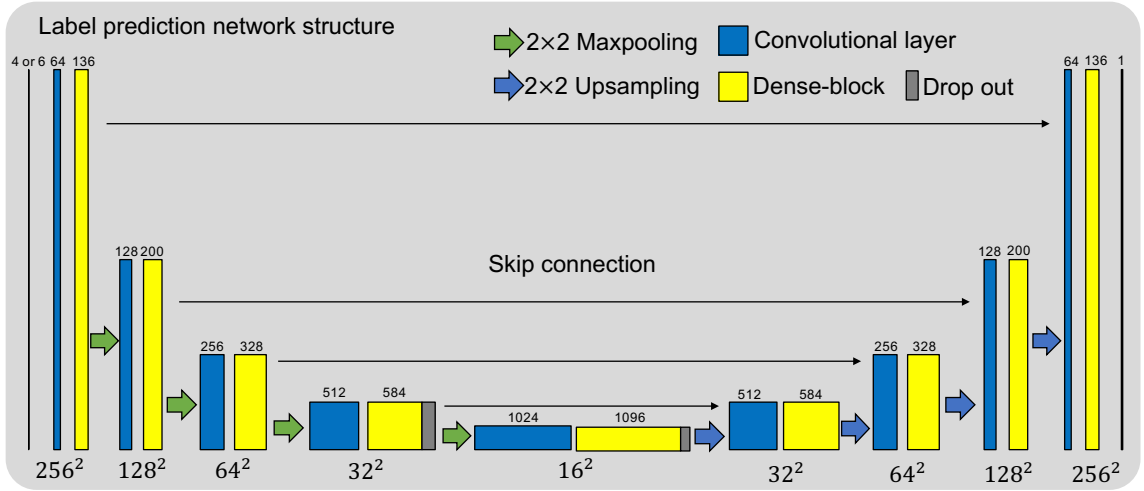


Figure 2-4: Neural network structure. Dense-blocks are added to facilitate efficient training. Skip connections are added to tunnel the high-frequency information from shallower layers to deeper layers with the same spatial scales. The lateral spatial size is marked by the number in the bottom of each block, and the layer depth size is marked by the number in the top of each block. The lateral sizes are compressed and decompressed by 2×2 maxpooling and upsampling layers, respectively.

reflectance image stack passes through the “encoder” path consisting of four dense blocks followed by the max-pooling layers, and the bottleneck feature maps are then fed into the “decoder” path with four dense blocks followed by upsampling layers. The skip connections bridge the lower-level activation maps with higher-level activation maps and preserve the high-frequency information. I use ReLU for all activation functions of the hidden layers and the sigmoid activation function in the final layer to form the final prediction ranging between 0 and 1. I use the negative PCC (NPCC) as the training loss [73]. I train the network using ADAM optimizer with 500 epochs and 0.1 training/validation splitting. No overfitting is observed during the training.

2.3.6 Quantitative evaluation of network prediction

Image pixel-level quantitative evaluation

I use the PCC to evaluate the performance of the regression-type of problems. Specifically, the PCC is used to quantify the prediction quality for pervasive subcellular

features, including DNA, endosome, actin, and Golgi apparatus labels. It computes the statistical correlation between the predicted and ground-truth IF image patches and is able to quantify the pixel-level similarity on the fine subcellular features. The PCC between the prediction X and the ground truth Y (each image is reshaped to a N dimensional vector) is computed as

$$\text{PCC} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2} + \epsilon} \quad (2.2)$$

where $\epsilon = 10^{-10}$ is a small regularizer to prevent zero denominator, \bar{X} and \bar{Y} denotes the mean, and i is the index of each vector. The value of the PCC ranges from -1 to +1, where ± 1 indicates total positive or negative correlation and 0 indicates no correlation. The PCC computation is implemented by a custom code in Python. The PCCs are computed on the testing image patches, each containing a 128×128 -pixel FOV. In total, the statistics from 676 patches from four large FOV testing images of size 908×908 pixels at each staining condition are aggregated and shown in the violin plots in Fig. 2-12B.

I use the AUC to quantify the detection performance for identifying the selective cell events including the proliferation and apoptosis. To plot the ROC of the detection label performance, the ground-truth labels are first binarized with certain thresholds. I use the well-established Otsu’s method to reliably compute the binarization threshold values [74], which finds the optimal values that maximize the intensity variance between the signals and the background based on a histogram clustering criterion. Specifically, the threshold values are calculated based on the aggregated histogram from the entire ground-truth dataset of proliferation and apoptosis, and are found to be 0.31 and 0.11, respectively. Next, each continuous-valued pixel in the predicted image is regarded as the predicted probability of expressing the IF label at this pixel.

By using the binarized ground-truth images as the target, the predictor (the trained CNN) achieves different pixel-wise True Positive Rate (TPR) and False Positive Rate (FPR) under different detection thresholds on the predicted images. By varying the detection thresholds, the TPR and FPR as functions of the thresholds can be plotted on the ROC curve. The AUC measures the area under the ROC curve and provides an aggregated quantification of the performance across all possible detection thresholds. The AUC is computed by the built-in functions ‘roc_curve’ and ‘auc’ in the scikit-learn module in Python. The AUCs are computed on the testing image patches, each contains a 128×128 -pixel FOV. In total, the statistics from 676 testing image patches from four large FOV testing images of size 908×908 pixels at each condition are aggregated and shown in the violin plots in Fig. 2-12B.

Image patch-level quantitative evaluation

To evaluate the spatial variations of the prediction performance, I construct the PCC map for each label prediction. To do so, 169 consecutive image patches are obtained from each large FOV image by cropping the image with a 128×128 -pixel sliding-window and 64×64 -pixel overlap between the neighboring patches. When computing the patch-wise PCC, I noticed that the lower accuracy regions generally align with the background region, where the local signal density and SNR is relatively low. This means that the patch-wise PCC computation is sensitive to low-SNR regions and PCC outliers in these regions should be eliminated. Accordingly, I use a median-based outlier detection and removal algorithm to remove these background outliers when constructing the violin plots. I added a median-based outlier detection and removal algorithm to remove the outliers and all the sub-FOV PCC shown on the violin plots in Fig. 2-12. The outlier detection and removal algorithm are implemented by a Matlab built-in function ‘filloutliers’ with findmethod of ‘median’ and fillmethod of ‘pchip’. The ‘median’ method finds the outliers as values that are three scaled

truth labeled nuclei are first segmented using ‘IdentifyPrimaryObjects’ module.

2) Predicted target IF images are segmented to find the predicted-positive (POS) cell nuclei. The POS are segmented using the same module as in 1).

3) The intersecting operation on GTP and POS provides the true positives (TP). I utilized ‘RelateObjects’ module to label POS with GTP as reference and ‘Filter-Objects’ module to filter out POS that does not contain GTP labels. This operation takes the intersection of POS and GTP, which are TP.

4) Taking the complement of TP from POS yields the false positives (FP). I label POS with TP as the reference and filter out POS that contains TP labels using the same modules.

5) The co-stained DNA channel ground truth are then used to segment all nuclei, denoted as ALL.

6) The complement of GTP from ALL is the ground truth negatives (GTN); the complement of POS from ALL is the predicted-negatives (NEG). I used the same complement operation on ALL with reference POS to get NEG and on ALL with reference GTP to get GTN.

7) Then I intersect NEG with GTN using the same operation as described in (3) to pick out true negatives TN.

8) False negatives FN are then obtained by taking the complement NEG with TN reference using the same operation as described in (4).

For proliferation fluorescence prediction, the intensity values are also aggregated within each type of cell mask regions to mark the scatter plot with individual detection performance labels.

Subsequently, I compute the cell-level detection metrics, including the sensitivity and specificity. The sensitivity (a.k.a recall) is computed from the TPR, and the specificity (a.k.a selectivity) is computed as the TNR as

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{TNR} &= \frac{\text{TN}}{\text{N}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \end{aligned} \tag{2.3}$$

2.3.7 Saliency map visualization

Our network performs image-to-image translation and can be treated as a function relating the input and output images. On the other hand, the norm of the network output is a scalar function. As a result, computing and visualizing the gradient over the norm is still possible. Based on this notion, I compute the saliency map as the gradient of the norm of the output with respect to the given input image. The practical implementation is achieved by the automatic differentiation feature in TensorFlow, as illustrated in Fig. 2·6.

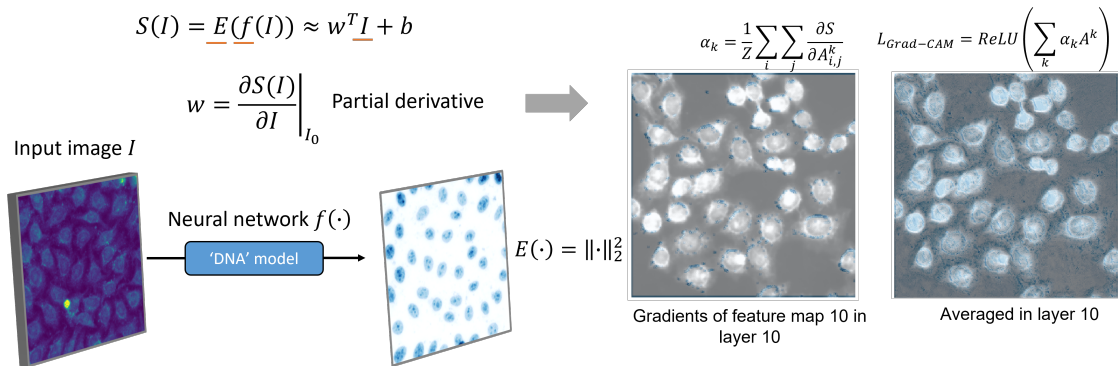


Figure 2·6: Saliency computation. The saliency map is determined by the gradient of the norm of the output with respect to the given input image. This shows an example of the calculated gradients of filter 10 in layer 10 and the averaged and smoothed gradient of all filters in layer 10 with respect to the input image I .

I utilized a built-in function ‘visualize_saliency’ in ‘vis.visualization’ module from an open-source packages ‘Keras-vis’ to enable the gradient computation. This is conveniently achieved by the following single-line script with several arguments to be

specified. `saliency_map= visualize_saliency(model, layer_idx=137, seed_input=seed, grad_modifier='absolute', backprop_modifier='guided')` Our output layer is specified as the last layer - layer 137 with the seed or input being the first layer of the network. I specified the gradient modifier as the absolute values of the gradient, which shows the regions in the input that contribute most to the change in the output regardless of the sign of the change (i.e. negative or positive). I also used the guided backpropagation to propagate only the positive gradients for positive activations to achieve a smoother visualization. The saliency map was computed for each network and on different sample input (for varying sample batches / fixation conditions). The inputs are 256×256 -pixel image stacks randomly selected from the testing groups under the six sample conditions. The computed saliency maps are normalized to have a uniform range between 0 and 1 for visualization.

2.4 Results

I show the individual fluorescence prediction results by both qualitative visualizations and quantitative evaluations on six models independently.

2.4.1 Qualitative fluorescence prediction results

To evaluate the performance of the DL models, I take measurements on fixed HeLa cells containing in-total six IF labels, including DNA (Hoechst), Golgi apparatus (GM130), endosome (EEA1), actin (Phalloidin), proliferation (EdU), and apoptosis (TUNEL). Specifically, five separate batches of IF staining are performed with GM130, EEA1, Phalloidin, EdU, TUNEL, each of which is co-stained with Hoechst. I then train six networks for performing individual IF label predictions using paired reflectance-fluorescence image dataset.

After training, I first evaluate each network's prediction accuracy on unseen reflectance input from the same cell batch. Fig. 2-7 shows the label-free input, the

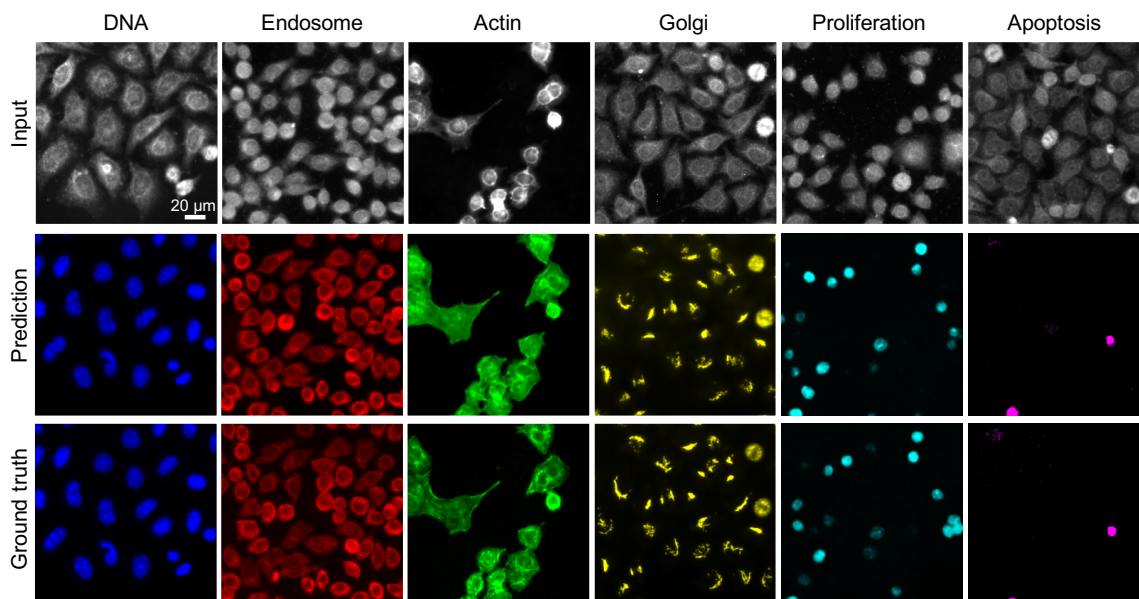


Figure 2-7: Visualization of the results from the six IF prediction networks. The rows show sample darkfield reflectance images from each input stack, the network’s IF prediction, and the ground-truth IF image, respectively. The columns show six IF labels covering nuclei (DNA), endosome, actin, Golgi apparatus, proliferation and apoptosis.

individual IF ground-truth, and the prediction for all six labels. The predicted sub-cellular structures and cell states have excellent visual agreement with the ground-truths. Characteristic morphological features are clearly recovered, including rounded nuclei, cytoplasmic endosome, spreading cell membrane (actin), and Golgi apparatus at the nuclei periphery. Selective cellular events or functions such as proliferation and apoptosis are also captured by the DL predictions. Additional examples of the prediction results are shown in Fig. 2-8, Fig. 2-9 and Fig. 2-10.

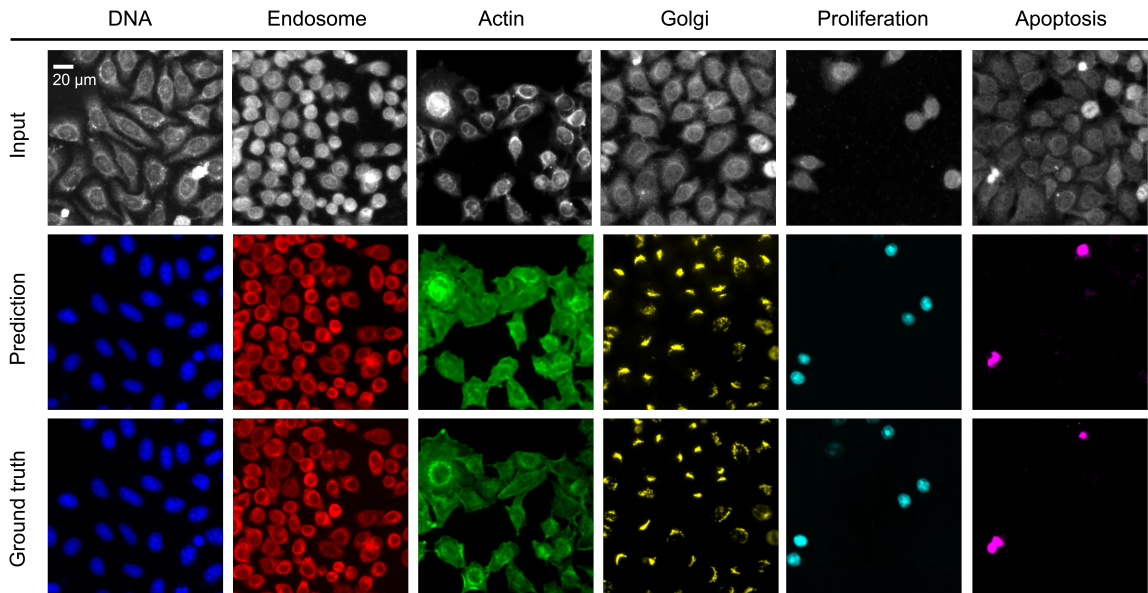


Figure 2-8: Additional individual prediction results. The rows show a sample reflectance image from each input stack, the network’s prediction, and the ground-truth IF image, respectively. The columns represent six IF labels including nuclei (DNA), endosome, actin, Golgi apparatus, proliferation and apoptosis.

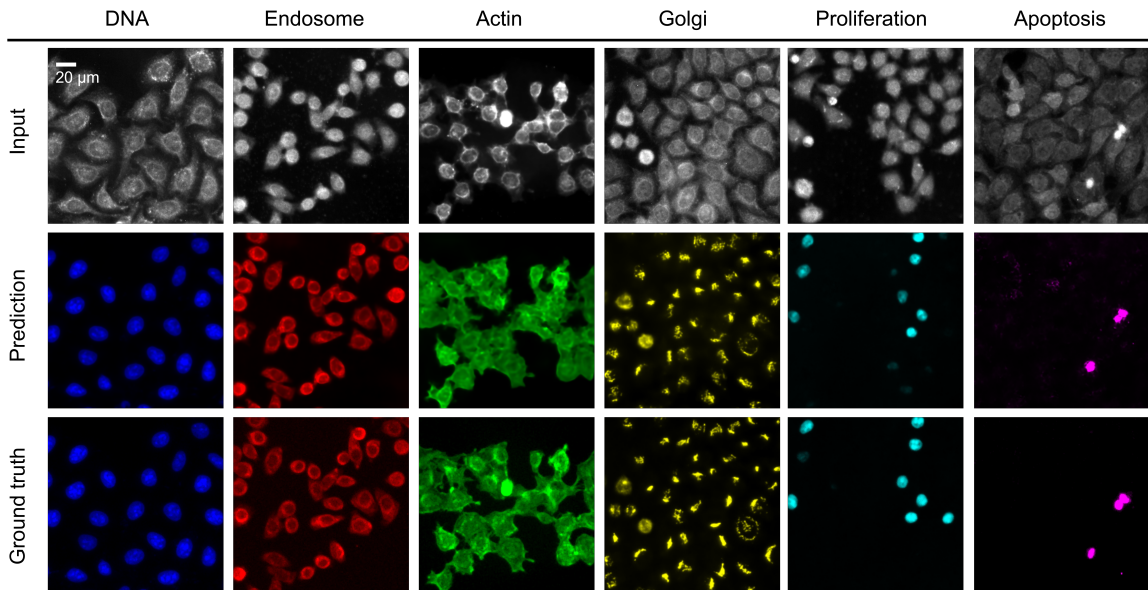


Figure 2-9: Additional results. Rows: sample reflectance image, network prediction, ground-truth IF image. Columns: IF labels including nuclei (DNA), endosome, actin, Golgi apparatus, proliferation and apoptosis.

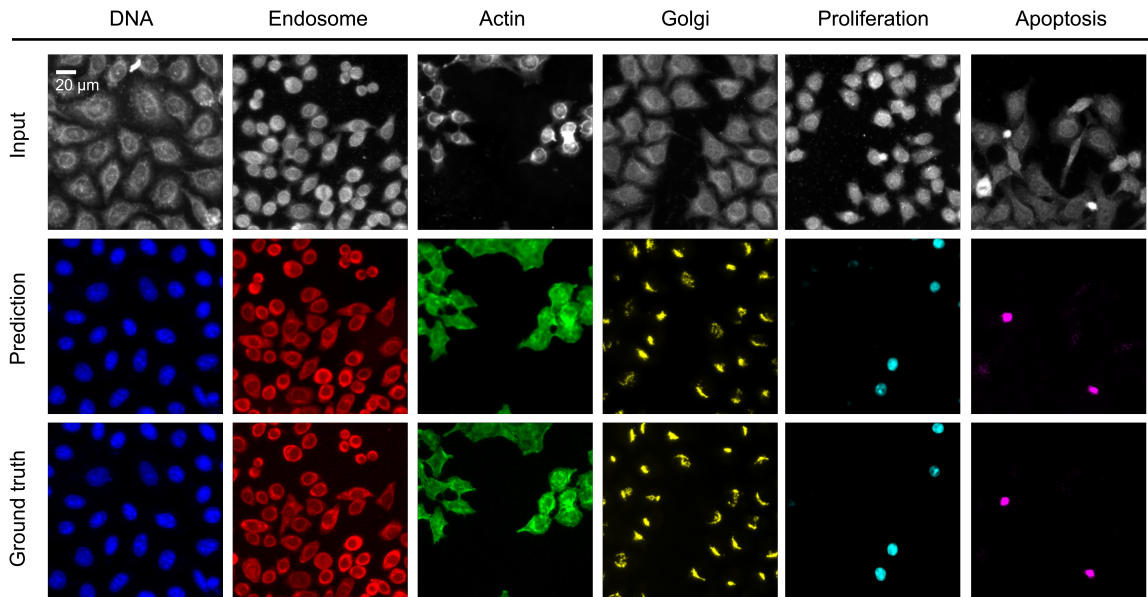


Figure 2·10: Additional individual prediction results.

A major goal of my study is to investigate the structural contrast captured by different label-free reflectance modes and understand how they impact the DL-based fluorescence predictions. To this end, I conduct ablation studies on the drDPC input in Fig. 2·11. I show example prediction results of six labelling networks trained with two different input settings: one with only four-channel darkfield reflectance input, the other with four-channel darkfield reflectance and two additional drDPC images as the input. The drDPC input helps preserving the fine details in the actin boundaries and improving true positive predictions in proliferation label, as shown in the examples marked by the red arrows.

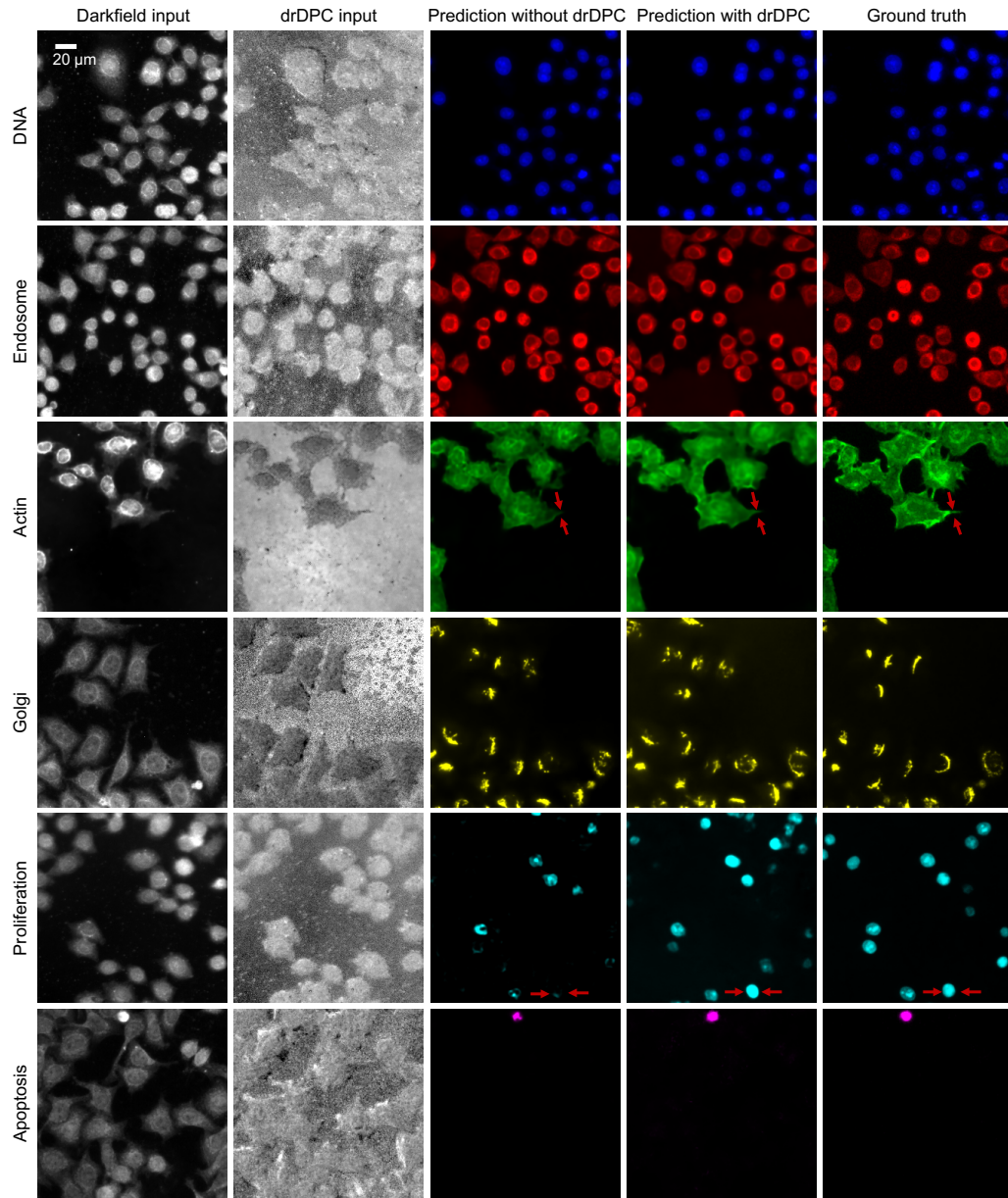


Figure 2.11: Network prediction with and without drDPC input channels. The first two columns show examples of the darkfield reflectance and drDPC images used for DNA, endosome, actin, Golgi, proliferation and apoptosis prediction, respectively. The last three columns show the predicted labels without the drDPC input, with the drDPC input and the ground truth labels, respectively.

2.4.2 Quantitative fluorescence prediction results achieve state-of-the-art

I quantify the prediction accuracy by computing several evaluation metrics on the network’s predictions based on the underlying cytometry tasks. Specifically, I formulate the predictions of the DNA, endosome, actin, and Golgi apparatus as regression problems because they are distinct subcellular structures and thus better described in their morphologies. Accordingly, the performance is quantified by calculating the image patch-wise Pearson correlation coefficient (PCC), which measures the pixel-level similarity of morphologies between the DL prediction and the ground truth. The distributions of the PCCs for the four IF label predictions are shown in the violin plots in Fig. 2-12A. The violin plots show the quantitative metrics for each IF label prediction. The upper and lower bounds of each grey bar represent the 25% and 75% quantiles, respectively; the center white point marked by the black dashed line denotes the median. In total, 676 testing image patches are aggregated for computing the statistics for each label. Notably, all four of the regression DL models achieve higher accuracy as compared to the current state-of-the-art techniques based on transmission label-free microscopy data [20]. The median PCCs on the DNA, endosome, actin, and Golgi apparatus label predictions achieve 87.25%, 91.85%, 92.01%, and 59.82%, respectively. These results agree well with the example visualizations in Fig. 2-7. The cellular features in the reflectance images associated with the corresponding fluorescence label are clearly visible.

I formulate the prediction of the proliferation and apoptosis as detection problems, which is more biologically meaningful because they are selective cell states. The proliferation labels are only present in the DNA-replicating cells. The apoptosis labels are only for cells undergoing programmed deaths. Accordingly, the performance is first quantified by the pixel-level Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) to measure the ability of separating the positives (i.e. those

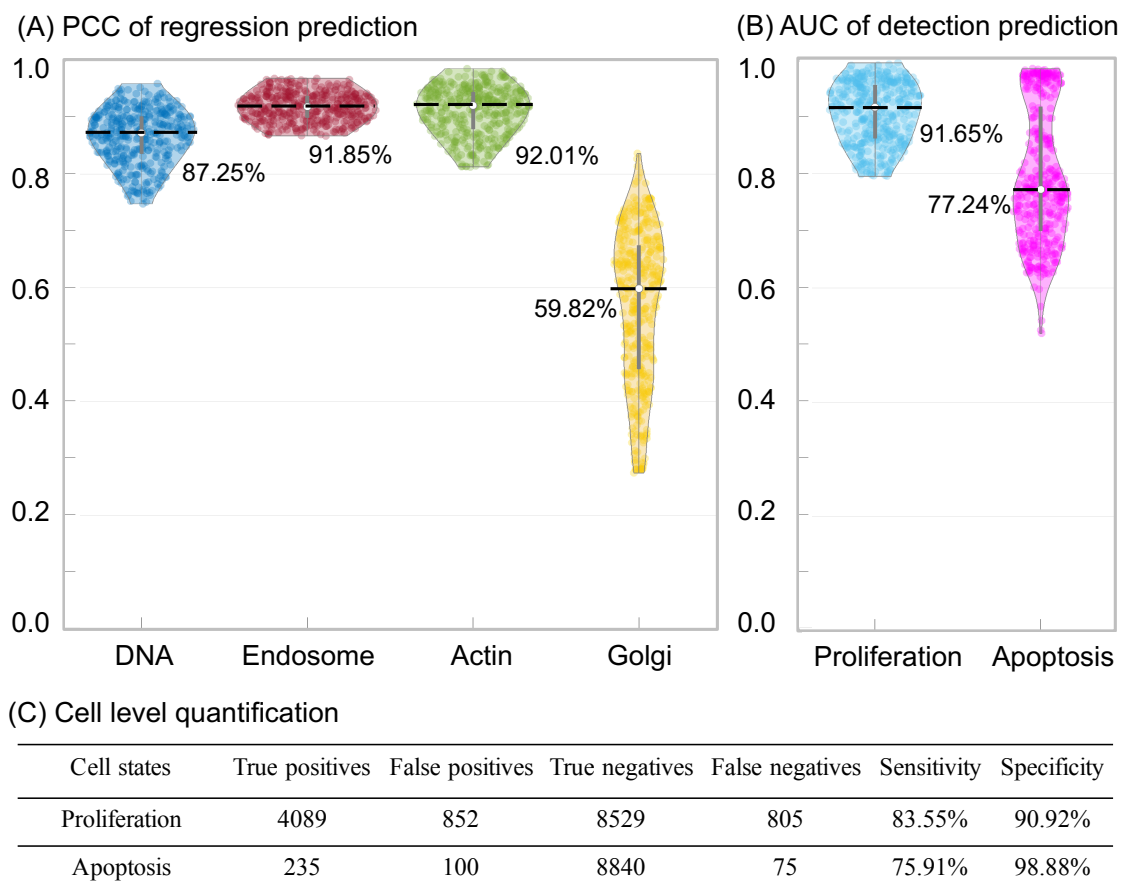


Figure 2-12: Quantitative evaluation of the DL prediction. (A) The image patch-wise PCCs of the predictions for nuclei (DNA), endosome, actin, and Golgi apparatus, which evaluates the pixel-level similarity between the regression-type predictions and ground-truth subcellular features. (B) The image patch-wise AUC of the proliferation and apoptosis predictions, which assess the pixel-level detection accuracy. (C) Quantitative evaluation of the cell-level detection performance of the proliferation and apoptosis predictions.

expressing the fluorescence) from the negatives at each pixel. The calculation is image batch-wise, the same way as I calculated PCC. The distributions of the AUCs over all batches for the two IF labels are shown in the violin plots in Fig. 2-12B. The median detection accuracies are 91.65% and 77.24% for the proliferation and apoptosis, respectively.

Next, I further evaluate the cell-level detection performance to give a more direct

assessment on these two label predictions. To do so, I perform single-cell segmentation on both the predicted and the ground-truth IF images and then identify each prediction as one of the four possible detection outcomes, including the true positive (TP), true negative (TN), false positive (FP), and false negative (FN), from which I compute the cell-level detection metrics, including the sensitivity and specificity. As summarized in Fig. 2-12C, the cell-level proliferation prediction achieves 83.55% sensitivity and 90.92% specificity; the apoptosis prediction achieves 75.91% sensitivity and 98.88% specificity. Notably, the scattering features in the proliferating cells cannot be easily perceived from the raw reflectance images, yet the DL model can capture the salient structural features with high accuracy. The balanced high sensitivity and specificity validates the reliability of the DL models for identifying these highly selective cell states/events.

I also quantitatively evaluate the effect of drDPC channels using the same metrics in Fig. 2-13. Quantitative metrics including PCC and AUC on 128 testing set images for six fluorescence predictions from models trained with two different input settings, including four-channel plain darkfield images of different illumination patterns or six-channel images consisting of the four-channel darkfield images and additional two-channel drDPC images. As expected, the additional drDPC channels substantially improve the actin IF label predictions by 9% since drDPC clearly highlights cell topography and boundaries as compared to the plain darkfield images. Notably, the additional drDPC channels also dramatically improve the prediction accuracy for proliferation and apoptosis IF labels, by 8% and 17%, respectively. I hypothesize that this is because distinct structural features exhibited during proliferation or apoptosis can be more prominently displayed by drDPC and subsequently recognized by the DL model.

Although these scattering signals are entangled with other signals in the raw label-

Fluorescence labels	DNA	Endosome	Actin	Golgi	Proliferation	Apoptosis
Accuracy without drDPC	87.25%	91.85%	83.25%	59.82%	83.90%	60.20%
Accuracy with drDPC	85.20%	90.07%	92.01%	57.20%	91.65%	77.24%

Figure 2-13: Quantification of network prediction with and without drDPC input channels. The prediction accuracy of DNA, endosome, Golgi degraded slightly with the additional drDPC input channels possibly because the drDPC morphological features do not add useful information for their predictions. The performance substantially improved for actin, proliferation and apoptosis with the additional drDPC input channels, indicating that the DPC morphological features contain rich information for their predictions.

free images, my result shows that the DL models are able to recognize and distill these salient features with high accuracy. As visualized in the violin plots and quantified by the 25% and 75% quantiles in Fig. 2-12A, different amounts of variations in the prediction accuracy are seen for the four labels. In general, I observe that the larger deviation of the prediction from the ground-truth, as measured by the median value, is also associated with larger accuracy variations. To investigate these spatial variations, I visualize the patch-wise PCC as a spatial map in Fig. 2-14. The prediction and ground truth images across 908×908 -pixel FOVs along with their patch-wise PCC map displayed on a 13×13 grid. Each PCC value is computed from a 128×128 -pixel image patch. The map is computed with a 64×64 -pixel overlap between the neighboring image patches. I show that the low PCC regions correspond spatially with the background region in the sample. Other than the background outliers, the PCCs are consistent across all the cell regions for all the label predictions.

Low-value PCC outliers are generally observed in background regions, and hence are removed by a standard algorithm. Other than the background outliers, the PCCs are consistent across all the cell regions for all the four IF label predictions, which demonstrates the overall robustness of the DL model.

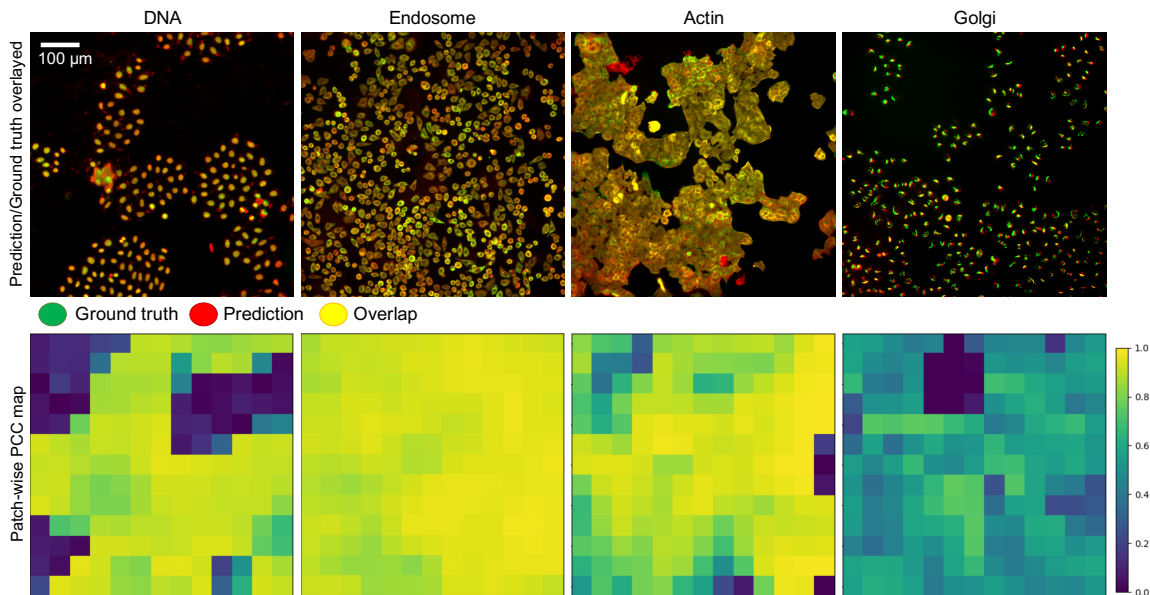


Figure 2.14: Spatial variation quantification of prediction accuracy. (Top row) Red: the prediction of four ‘regression’ labels, including the DNA, endosome, actin, Golgi apparatus; Green: the ground truth overlay. (Bottom row) the corresponding PCC maps.

Overall, these individual label prediction results validate my hypothesis that the improved sensitivity and resolution in reflectance images contain rich morphological features that can be utilized effectively for structural phenotyping by DL.

2.5 Neural network saliency analysis

Deep neural networks have shown high expressivity for complex models, but suffer from poor explainability. Many theoretical explanations for the DL model have resorted to statistical perspective while treating the overall model as a “black box”. Instead, I utilize the “attention”-based technique [71] to elucidate on the specific label-free subcellular features that contribute to the fluorescence prediction. To do so, I treat the trained DL model as a mapping function between the input and the output. I then visualize the network’s gradient with respect to the input and extract the salient features (i.e. those having the largest gradients) the network pays most

attention to. The resulting “saliency map” highlights the most important features contributing to the IF prediction. By doing so, the saliency map directly evaluates the specificity of the structural features extracted from the reflectance images and how they are transformed to the target fluorescence labels by the network.

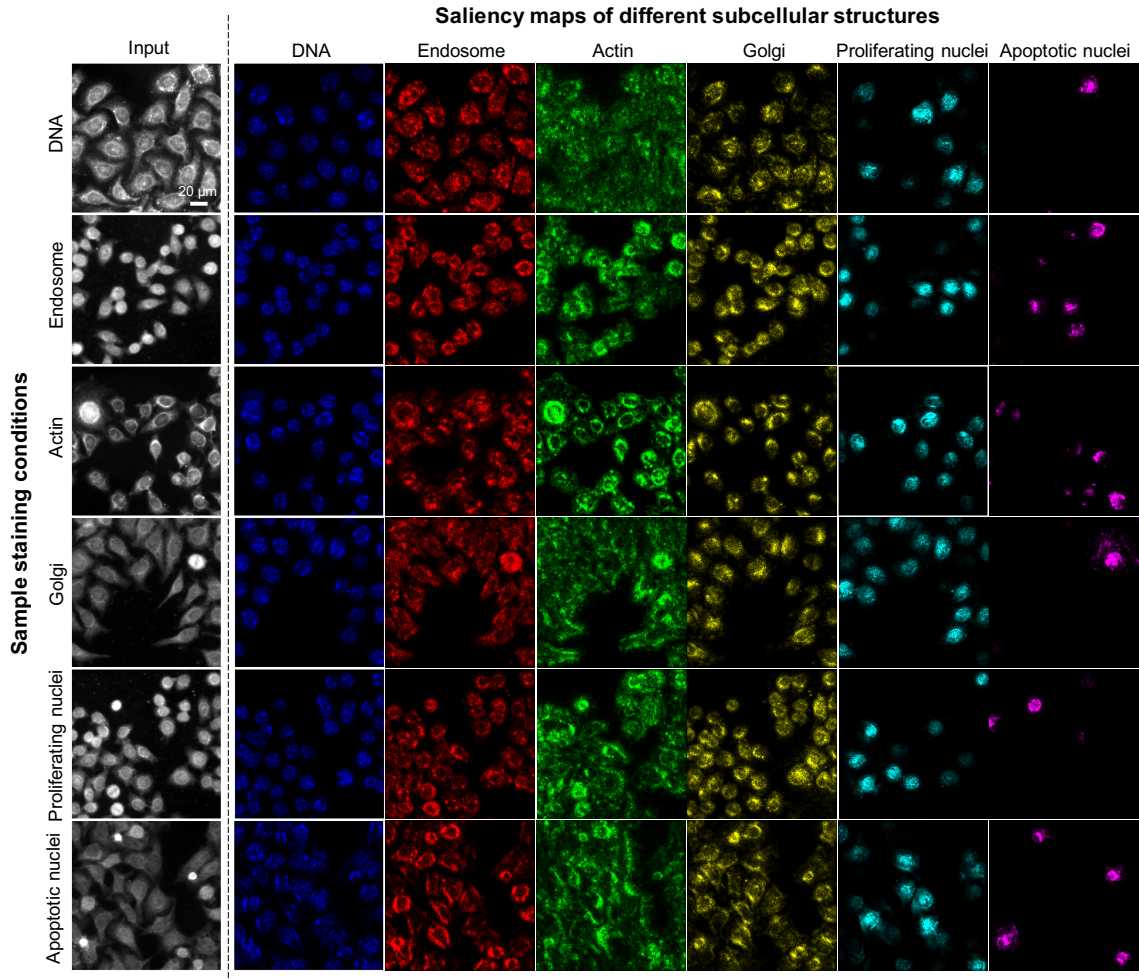


Figure 2-15: Saliency maps from each network across six cell batches with different staining conditions. The columns show the label-free input (the first darkfield reflectance channel) and the saliency maps for six different IF labels, including DNA (blue), endosome (red), actin (green), Golgi apparatus (yellow), proliferation (cyan), and apoptosis (magenta). The rows show the label-free input and the saliency maps from six cell batches under different staining conditions. The saliency maps show good consistency across different batches and highlight distinct morphological features.

Fig. 2.15 shows the computed saliency maps for each network across different sample batches and labeling conditions. Importantly, distinct subcellular features not only are highlighted by the network’s saliency map, but also have good visual correspondence to the targeted fluorescence label. By inspecting different columns, I show morphologically distinct features from different networks, indicating that different networks can indeed learn to recognize and focus on specific features present in the label-free images. For example, the DNA saliency maps show emphasis on nuclear boundaries and some subcellular structures. The actin saliency maps show concentration over the whole cell and spreading out to the membrane boundaries. The saliency maps for Golgi apparatus generally form shapes in “partial-moon” or circular lines. By contrast, the saliency maps for proliferation and apoptosis show that the network selectively pays attention to certain features around the nuclei. In addition, the saliency maps show that my network learns to extract invariant structural features specific to the underlying fluorescence label regardless of the cell preparation and labeling processes. Across different rows, I observe consistent saliency maps under different sample batches / staining conditions for the same labeling network.

2.6 Summary

In this study, I have presented a DL-augmented label-free cytometry technique that accurately predicted six fluorescence targets in-parallel at the single-cell level. The accuracy has been improved up to $3 \times$ in predicting subcellular structures as compared to the current state-of-the-art. Remarkably, the DL model is able to accurately recognize subcellular organelles, such as Golgi apparatus reconfigure during the cycle of proliferation, as well as to distinguish subtle morphological differences between the proliferating and non-proliferating cells. These results demonstrate the data-driven model’s unique capability of holistically extracting “non-intuitive” structural

features from the label-free imaging data on a large cell population. The specificity for predicting cellular features by the DL model is illuminated by the saliency maps. This analysis demonstrated the ability of the DL models in processing highly complex and entangled structural information from scattering images.

Chapter 3

Digital Multiplexed Fluorescence Labeling for Single-cell Cytometry

3.1 Background

Cell morphology features are powerful phenotypical readouts, which have been the basis for pathology for decades. They are also the underlying mechanisms for varieties of imaging cytometry and high-content screening platforms to characterize pathological changes and responses to drug treatments [76]. The most widely used approach for imaging readouts are fluorescence labels that highlight specific subcellular components or cell functions through immunofluorescence (IF), fluorescent reporter cells or dyes. However, the throughput of these approaches is fundamentally limited by the physical process of labeling. The IF staining is labor-intensive and generally requires cell fixation that does not allow kinetic observations of live cells over time. Fluorescent reporter cells are permissive for longitudinal live cell imaging. However, the process of gene editing and validation takes a significant amount of time and can be difficult to introduce multiple markers within the same cells for multiplexed analysis. Regardless of the fluorescence labeling approaches, the overlapping of the fluorescence emission spectra further limits the multiplexing capability. To alleviate these limitations, here I develop a label-free single-cell cytometry that is highly multiplexed and can forgo the physical staining via a DL-augmented digital labeling method.

3.2 Overview

Following my work presented in Chapter 2, I further applied my developed technique in single-cell cytometry and profile analysis to showcase the potential of digital fluorescence labeling in the field of cell imaging and image analysis.

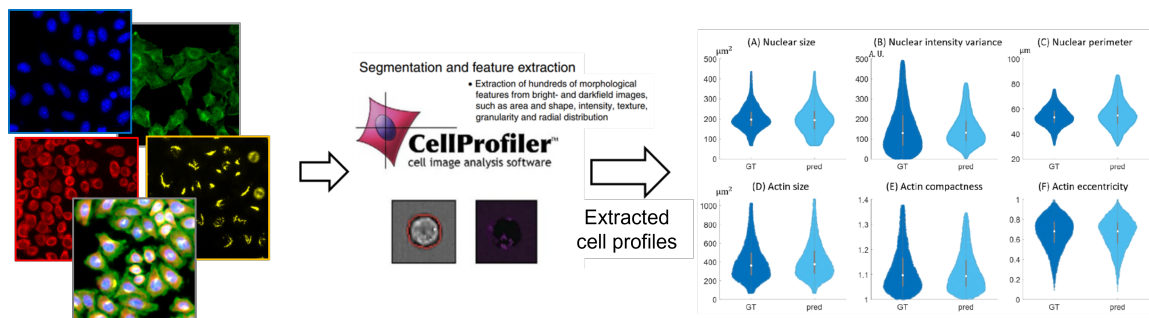


Figure 3-1: Overview of the cell cytometry and profile analysis based on multiplexed digital fluorescence labeling. I evaluate several cellular features on the digitally multiplexed readouts using cell profiler. Several cellular metrics are extracted from the predicted IF labels and from the ground truth, and visualized in the violin plots.

One distinct contribution of this work is to advance beyond the prediction of fluorescence images, and to demonstrate accurate structural phenotyping and quantitative single-cell cytometry using digitally multiplexed fluorescence images. Importantly, I show that the DL model can correctly capture and predict characteristic subcellular features during the cell cycle, including morphological changes of nuclei and Golgi apparatus. I also show that my DL model can capture the structural features of cell proliferation and recapitulate the DNA duplication through the cell cycle. Remarkably, the label-free structural features identified in proliferating cells are not obvious by visual inspection of the raw images, demonstrating that my holistic DL model can potentially capture novel cellular attributes with high accuracy. Another distinct advantage of multiplexing several fluorescence markers is that it enables the development of multiple quantitative metrics for imaging cytometry and phenotyping across the large cell population and at the single-cell level (e.g. cell

size, nuclear-cytoplasmic ratio, nuclear roughness, Golgi eccentricity, etc.). As other “omic” platform technologies are rapidly being developed to evaluate biological processes at various levels (e.g. genomics, transcriptomics, proteomics, metabolomics), single-cell level structural metrics will complement these population-based studies, particularly when a contextual phenotypic shift is expected only for a subset of cells. As a demonstration, I evaluate several cellular features, including morphology and fluorescence expressing intensity on the DL predicted digitally multiplexed readouts, as shown in Fig. 3.1.

3.3 Methods

3.3.1 Digital cytometry analysis

I develop a digital cytometry analysis framework for exploring the interdependencies of different fluorescence markers on the multiplexed predictions. A commonly used flow cytometry analysis is performed by displaying the scatter plot of single-cell level proliferating DNA concentration in the log-scale against the DNA concentration in the linear-scale. Different from flow cytometry that directly collects the integrated fluorescence intensity from each cell, my method performs imaging with subcellular resolution across a large cell population. As a result, I first perform cell segmentation and then aggregation of the fluorescence signals within each cell region to carry out the single-cell digital cytometry analysis. I perform the digital cytometry analysis to relate the cell-level Hoechst and EdU fluorescence concentration using CellProfiler [75]. Since our data contain co-registered two-channel fluorescence images with co-stained Hoechst and EdU, I can directly compare the ground-truth cytometry scatter plot with that from my multiplexed prediction. To generate the scatter plots, I process the co-registered Hoechst and EdU images and the multiplexed predictions using CellProfiler with a standard single-cell segmentation-based pipeline and extract

the paired cell-level fluorescence intensity data. A small value (10^{-10}) is added to the fluorescence intensity of the proliferating DNA before taking the log operation to avoid the singularity at 0. Three distinct clusters representing the S, G1, and G2/M phase are clearly shown in the digital cytometry scatter plots.

To generate the cytometry scatter plots, I feed the co-registered Hoechst and EdU images to CellProfiler and then perform the following steps as shown in Fig. 3-2.

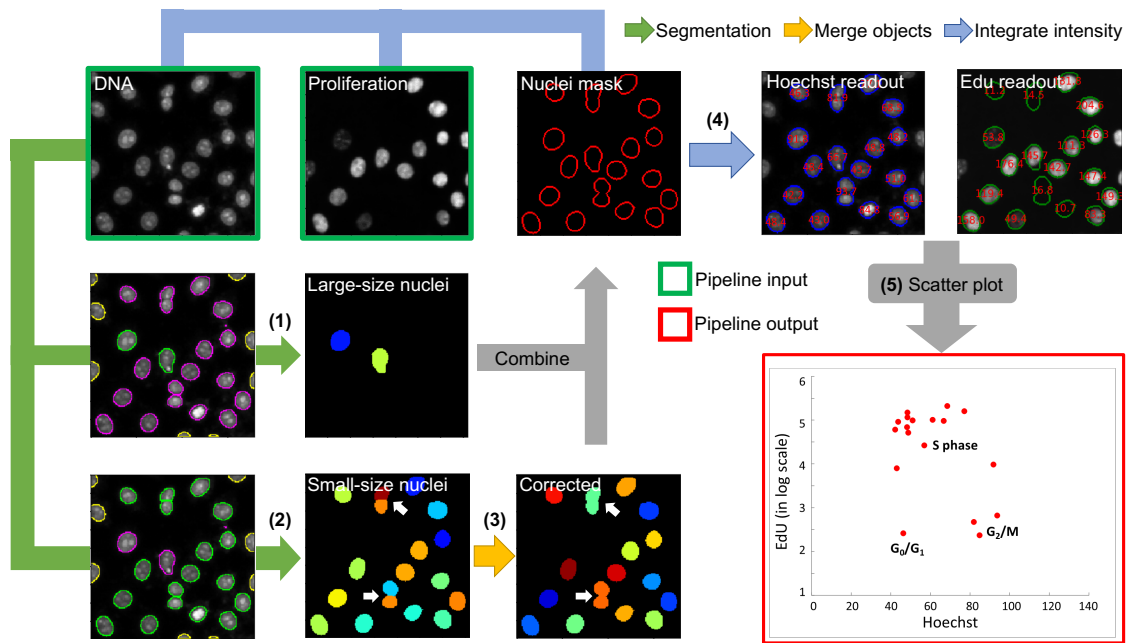


Figure 3-2: Procedure of digital cytometry analysis. Image analysis pipeline for digital cytometry scatter plot. The pipeline applies to the DNA/proliferation image pairs from both the ground truth and the multiplexed predictions. The segmentation, ‘merge objects’ and ‘integrate intensity’ operations are implemented by the modules in CellProfiler.

I segment all the nuclei in the Hoechst image by the ‘IdentifyPrimaryObjects’ module and extract all the corresponding nuclei masks in the data. I then apply the masks to both Hoechst and EdU images and calculate the integrated fluorescence intensity inside the mask by the ‘MeasureObjectIntensity’ module. One confounding factor of this pipeline I find is that using the pure image data can mis-detect certain

M-phase cells during the segmentation. Specifically, in anaphase and telophase (sub-phases in M phase) when a cell is about to split into two, the typical morphology is that a cell is attaching or intimately close to a nearby one. While common flow cytometry readout treats the two cells as one (since they are likely to be flown through the fluidic channel simultaneously) and gives the correct readings of DNA and proliferating DNA expression, my digital cytometry does not with a plain segmentation algorithm in CellProfiler. To overcome this issue, I refine the segmentation procedure by the following steps.

- 1) The DNA images are segmented to locate all the large-size nuclei using a large size constraint. I set the segmenting size to 24-35 pixels to find normal-sized nuclei that are prevalent in S and G1/G2 phases.

- 2) The same DNA images are then segmented to find all the small-size nuclei with a smaller size constraint. I set the segmenting size to 16-23 pixels to find small-sized nuclei that exist in the M phase.

- 3) An additional correction operation is performed on small-size nuclei to merge the incorrectly separated cell segments (those during mitosis) into a single segment. The corrected small-size nuclei masks are then combined with the large-size masks to provide a union of all the nuclei masks. I use the ‘MergeObjects’ module to combine the small-sized nuclei masks that are within a distance of 2 pixels into a single mask.

- 4) The DNA and proliferation image pairs are then masked. The intensity values within each mask are aggregated to provide the single cell-level concentration readout. The DNA channel corresponds to the Hoechst-stain concentration and the proliferation channel correspond to the Edu-stain concentration. I combine the merged masks and the normal-sized nuclei masks for computing the cell-level fluorescence intensities. I find this procedure successfully reduces the misdetection of M-phase nuclei and provides a more accurate scatter plot.

5) All the single-cell level two-channel readouts are plotted cell by cell. The paired cell-level fluorescence intensity data are plotted in Matlab. A small value (10^{-10}) is added to the fluorescence intensity of the proliferating DNA before taking the log operation to avoid singularity at 0. For the multiplexed predictions, I apply the same pipeline on the predictions from the DNA and proliferation networks. Three distinct clusters representing the S, G1, and G2/M phase are clearly shown in the digital cytometry scatter plots.

I construct an image processing pipeline to quantify the single-cell-level detection performance on the proliferation and apoptosis labels (denoted as the target IF in this paragraph). First, for each of the cell event labels, I perform segmentation on the co-stained DNA channel ground truth images to find all the nuclear localizations. Second, I segment all the nuclei in both the target IF prediction and ground-truth images. Third, each nucleus is labeled as one of the four possible detection outcomes. Specifically, a prediction is TP if both the prediction and the ground truth express the target IF. A prediction is TN if neither the prediction nor the ground truth express the target. A prediction is FP if the prediction expresses the target IF while the ground truth does not. A prediction is FN if the prediction does not express the target IF while the ground truth does. To further investigate the effect of false detections (FP/FN) on the prediction digital cytometry scatter plot, I apply the above pipeline and label each prediction as TP, TN, FP, or FN.

3.3.2 Cell profile analysis

I use CellProfiler [75] to generate the single-cell profiles across each fluorescence image. I feed the ground truth and the predicted IF images of DNA, actin, endosome, and Golgi apparatus to CellProfiler. After initial cell segmentation, single-cell level parameters of morphology and intensity distribution are computed automatically by different measurement modules in CellProfiler, including the fluorescence marker

size (area), compactness, eccentricity, fluorescence concentration, and single-cell-level fluorescence variance and contrast. In addition, I compute the compound metric, nuclear-cytoplasmic ratio (NCR), based on the multiplexed DNA and actin fluorescence labels. To compute the NCR, co-registered ground-truth / prediction images containing the DNA and actin labels are processed individually in CellProfiler. The NCR is computed as the ratio between the area of segmented nuclei and actin masks. The single-cell profiles also present the same outliers from the background regions, which are eliminated by the outlier removal algorithm when constructing the violin plots.

The single-cell profiles are analyzed by CellProfiler across each entire fluorescence image. I feed the ground truth and the predicted IF images of DNA, actin, endosome and Golgi apparatus to CellProfiler. After initial cell segmentation, single-cell level parameters of morphology and intensity distribution are computed automatically by different measurement modules in CellProfiler. Specifically, I use the ‘MeasureObjectSizeShape’ module to measure the fluorescence marker size (area), compactness, and eccentricity, the ‘MeasureObjectIntensity’ module to measure the fluorescence concentration, the ‘MeasureTexture’ module to measure the single-cell-level fluorescence variance and contrast. The fluorescence marker size is measured by the number of pixels in the segmented cell region. The compactness is computed by the mean squared distance of the cell mask pixels from the centroid divided by the area. The eccentricity is defined by the ratio of the distance between the foci of the effective ellipse that has the same second-moments as the segmented region and its major axis length. The concentration is computed as the sum of the intensities within the segmented masks. The contrast is measured by the local variation in a cell region. The variance is measured by the variation of the intensity values. The cell-level parameters are extracted from CellProfiler and then imported to Matlab to generate the

violin plots. NCR is a compound metric that involves multiplexed DNA and actin fluorescence labels. To compute this metric, co-registered ground-truth / prediction images containing the DNA and actin labels are processed individually in CellProfiler. NCR is computed as the ratio between the area of segmented nuclei and actin masks.

3.4 Results

I present visualized results of my multiplexed fluorescence labeling framework and cell cytometry and profile analysis.

3.4.1 Multiplexed prediction recovers biological accurate cellular structures

I demonstrate the digital multiplexing capability by feeding the same reflectance input to each network and make six different IF predictions in parallel. By doing so, multiple subcellular structures and cell states are revealed simultaneously. In Fig. 3-3A, the image multiplexes the nucleus, Golgi apparatus, actin, and endosome virtual IF labels in a single wide field-of-view (FOV) for a large cell population. In Fig. 3-3B, the virtual labels for proliferating and apoptotic cells are multiplexed with the darkfield reflectance input in the same FOV as in Fig. 3-3A. These multiplexed predictions are performed on the cell batch under the Golgi apparatus staining condition.

Importantly, my results show that the DL model can correctly capture and predict characteristic subcellular features during the cell cycle. During interphase, the nuclei have a regular rounded shape with nucleoli present, and Golgi apparatus is anchored primarily to one side of the nuclei (Fig. 3-3C-E). At this stage, cells that have initiated or ongoing DNA/chromatin replications have a positive signal for proliferation (Fig. 3-3F). When cells enter mitosis, the chromosomes start to condense toward the centers of the cells, and nucleoli disappears. Golgi apparatus undergoes vesiculation and fragmentation, and its components are found scattered throughout the

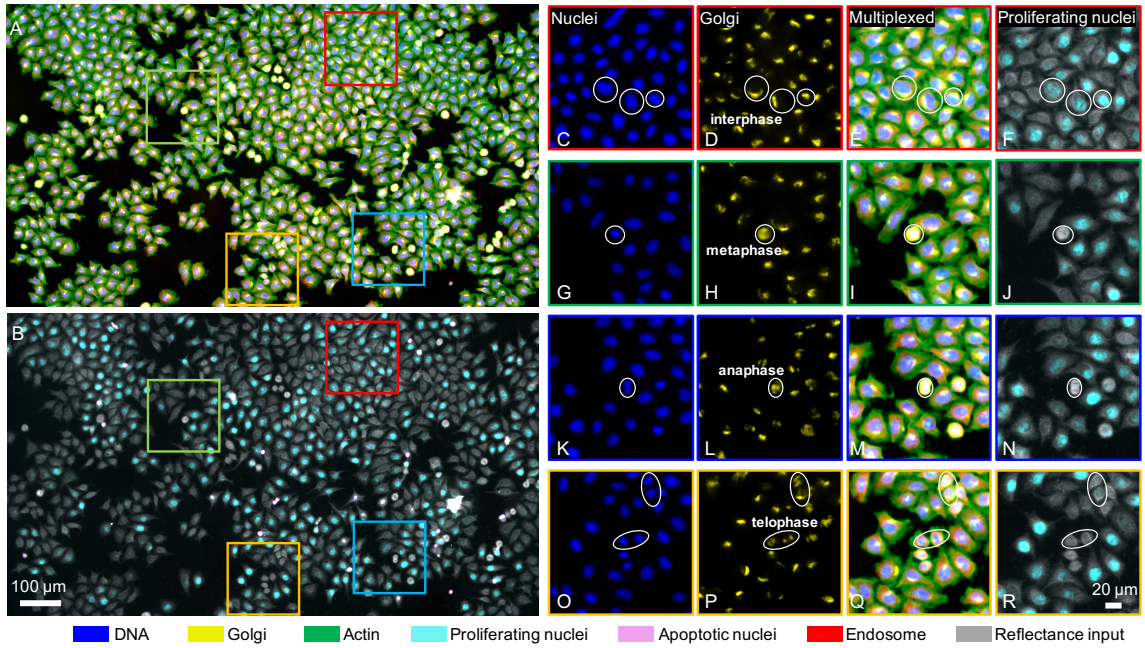


Figure 3-3: Multiplexed prediction on six IF labels from the same label-free input. (A) Visualization of the Full-FOV multiplexed prediction including DNA (blue), endosome (red), actin (green), and Golgi apparatus (yellow), and (B) proliferation (cyan) and apoptosis (magenta) from the same reflectance input (grayscale). (C-R) Zoomed-in of DNA, Golgi apparatus, multiplexed, and proliferation predictions. White circles indicate representative cell morphology during different phases of the cell cycle, including (C-F) interphase, (G-J) metaphase, (K-N) anaphase, and (O-R) telophase.

cytoplasm in the form of tiny (about 50-nm) vesicles, often referred to as the “Golgi haze” [77]. During metaphase, chromosomes align at the metaphase plate and the cell shape also changes dramatically, bulging into a sphere (Fig. 3-3G-I). Golgi haze appears rounded, with a shaded center where chromosomes are located (Fig. 3-3H). During anaphase, the duplicated chromosomes separate from one another and move to opposite poles of the spindle (Fig. 3-3K-M). During telophase, chromosomes start to de-condense, and begin to take on a more interphase-like shape (Fig. 3-3O-Q). In this stage, Golgi apparatus has also completed replication, and reassembled into two closely-spaced cell bodies, referred to as “Golgi twins” [77] (Fig. 3-3P). There is no DNA replication during mitosis, so the markers for proliferation (incorporation of the

fluorescent nucleoside, EdU) are absent for the metaphase, anaphase, and telophase (Figs. 3-3J, N, R). As shown in Fig. 3-3C-R, these structural, subcellular, cell cycle-dependent features are accurately captured and predicted by the DL model, which validates my hypothesis that label-free reflectance imaging and DL enable structural phenotyping.

To further demonstrate the robustness of this digital multiplexing procedure, I show additional examples of multiplexed predictions performed on different cell batches/staining conditions in Fig.3-4 and Fig.3-5.

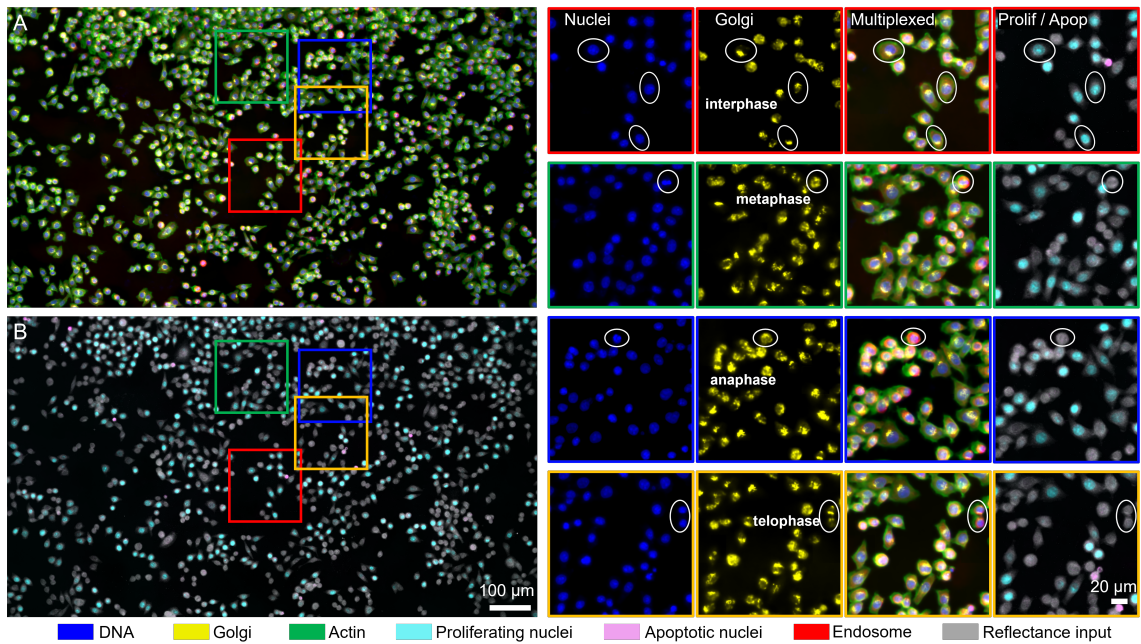


Figure 3-4: Additional multiplexed prediction results from the endosome staining cell batch. (A) Visualization of the Full-FOV multiplexed prediction including DNA, endosome, actin, and Golgi apparatus. (B) Proliferation and apoptosis from the same reflectance input. The right-hand-side are zoomed-in of DNA, Golgi apparatus, multiplexed, and proliferation predictions. White circles indicate representative cell morphology during different phases of cell cycle.

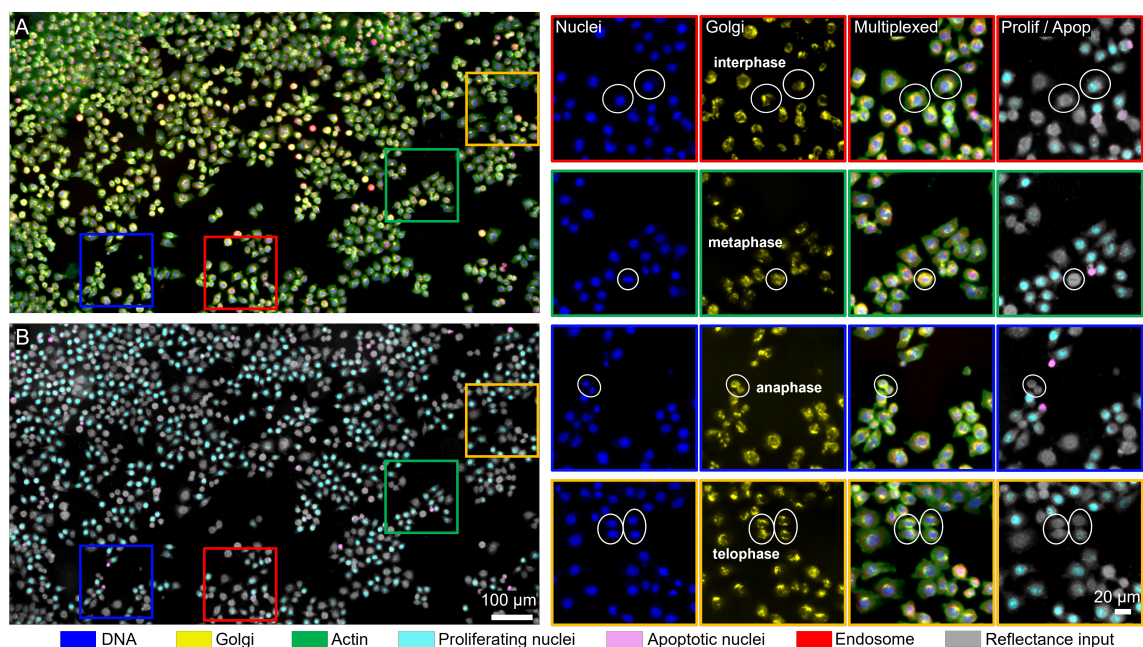


Figure 3-5: Additional multiplexed prediction results from the proliferation staining cell batch.

3.4.2 Cell profile analysis on multiplexed images allows phenotyping and quantitative cytometry

A distinct advantage of multiplexing several markers is that it enables the development of multi-variant quantitative metrics for imaging cytometry and phenotyping across the large cell population and at the single-cell level. As a demonstration, I evaluate several cellular features, including cell morphology and fluorescence intensity, on the digitally multiplexed readouts.

First, I generate fluorescence intensity scatter plots similar to those used in the flow cytometry, of EdU vs. Hoechst for all cells from the ground-truth and digitally multiplexed IF images in Fig. 3-6B and C, respectively. Evaluating the scatter plots on the population level, the DL-multiplexed prediction matches well with the ground truth, both of which show the increase in EdU and doubling of Hoechst intensity in the S and G2/M phases of the cell cycle, respectively. Next, I further evaluate

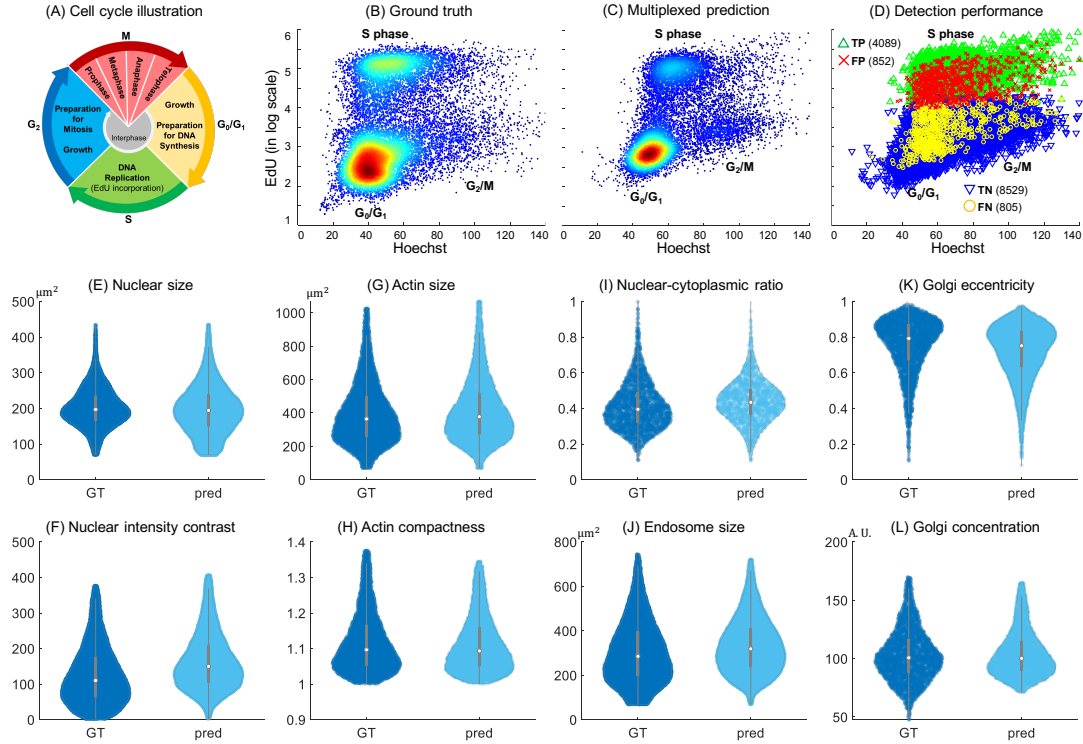


Figure 3-6: Cell profile analysis on digital multiplexed IF staining. (A) Illustration of the cell cycle. (B-D) The scatter plots for the whole cell-level EdU (proliferating DNA) and Hoechst (DNA) concentrations from (B) the co-stained ground-truth, (C) the DL-prediction, and (D) the detection performance for the EdU predictions quantified by TP, TN, FP, and FN, across the entire cell population under the proliferation staining condition. (E-L) The comparisons of the statistics of eight single-cell profile metrics extracted from the entire cell population in the ground truth (GT) and the DL-predictions (pred), including (E) nuclear size, (F) DNA (nuclear) fluorescence intensity contrast, (G) cell (actin) size, (H) compactness of the actin, (I) NCR measured by the area ratio between the nuclei and actin, (J) endosome size, (K) eccentricity of the Golgi apparatus distribution, and (L) concentration of the Golgi apparatus.

the results on the individual cell level by overlaying the detection outcome for proliferation of every cell onto the same scatter plot in Fig. 3-6D. The total numbers of sample cells collected from the ground truth and DL-predictions are 14778 and 14275, respectively. The numbers of TP, FP, TN and FN in (D) are 4089, 852, 8529, and 805, respectively. Specifically, by comparing the predicted proliferation IF label and the ground truth, each data point is labeled as one of the four detection outcomes

(TP, TN, FP, or FN). My analysis shows that the incorrect predictions (including FP and FN) tend to cluster around the boundaries between the S phase and S1 phases, leading to confusions in the DL predictions. There are relatively fewer incorrect predictions in G2/M phase, which are expected because of the distinctive morphological features during mitosis that can be easily captured by the DL model (e.g. metaphase, anaphase, and telophase in Fig. 3-3).

In Figs. 3-6E-L, I extract several biologically relevant single-cell profile metrics using the predicted IF labels and compare them with the ground truths, as visualized in the violin plots. In particular, I show the statistics of eight different morphological and subcellular structural parameters. The total numbers of sample cells collected from the ground truth and DL-predictions are respectively 20021 and 25257 in (E) and (F), 6183 and 5151 in (G) and (H), 2246 and 1491 in (I), 15944 and 22408 in (J), 3380 and 9392 in (K), 3380 and 3380 in (L). In Fig. 3-6E-F, I gather statistical data about the DNA label to measure the nuclear size and intensity contrast. In Fig. 3-6G-H, I evaluate the actin size (i.e. cell size) and its compactness. In Fig. 3-6I, I compute the nuclear-cytoplasmic ratio (NCR), an important marker for cancers, as the area ratio between the nucleus and actin. In Fig. 3-6J, I measure the endosome size. In Fig. 3-6K-L, I collect morphological parameters about the Golgi apparatus, including the eccentricity and concentration. All the single-cell profile metrics show good agreements between the predictions and the ground truths.

Additional metrics are provided in Fig. 3-7. The total numbers of sample cells collected from ground truth and DL-predictions are respectively: 20021 and 25257 in plots (A) (B) (C), 6183 and 5151 in plots (D) (E) (F), 15944 and 22408 in plot (G) (H) (I), 3380 and 9392 in plot (J) (K) (L). For all these single-cell profile metrics, the prediction and the ground truth show excellent agreement. These results clearly demonstrate that the DL-augmented label-free cytometry can provide comprehensive

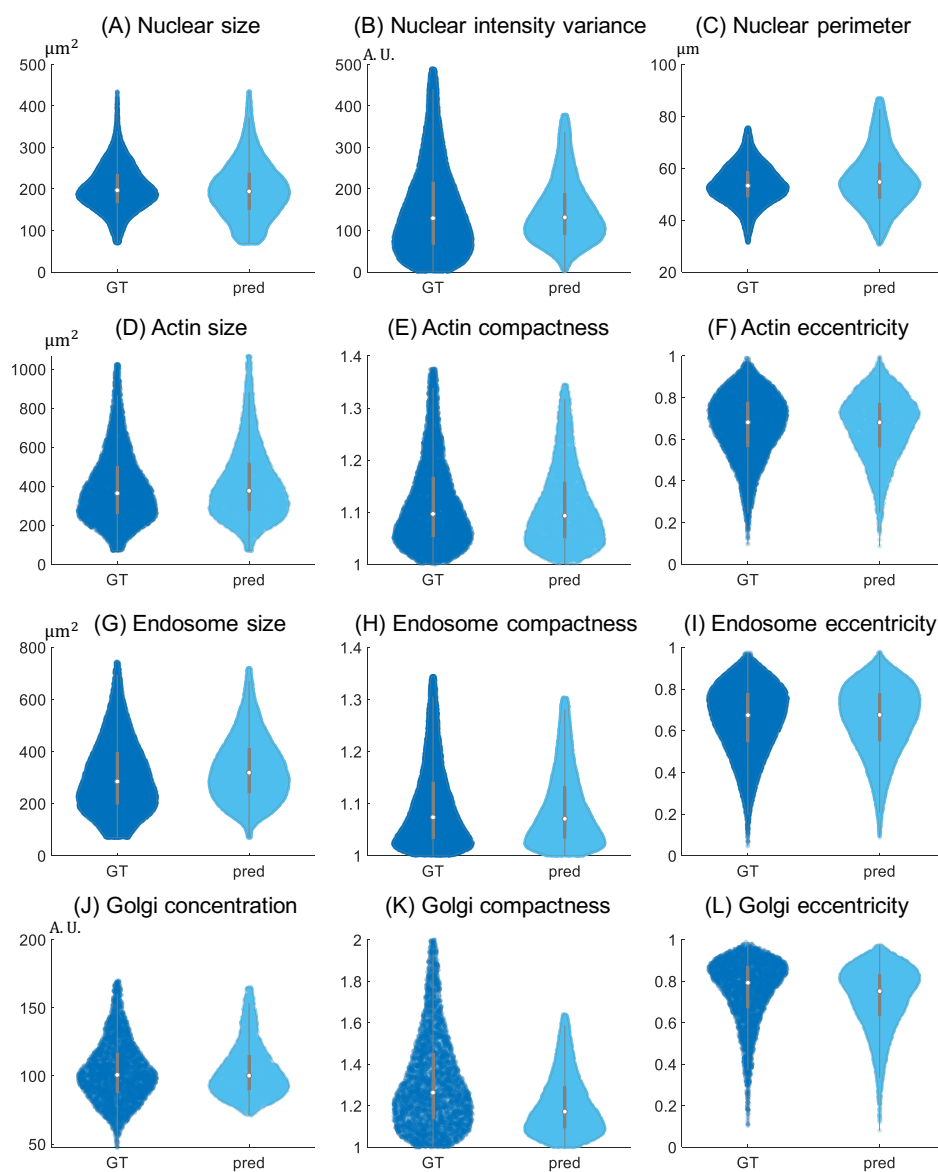


Figure 3-7: Additional single-cell profile metrics. The comparisons of the statistics of twelve different single-cell profile metrics extracted from the entire cell population in the ground truth (GT) and DL-predictions (pred), including (A) nuclear size, (B) DNA (nuclear) fluorescence intensity variance, (C) nuclear perimeter, (D) cell (actin) size, (E) compactness of actin, (F) eccentricity of actin, (G) endosome size, (H) compactness of endosome, (I) eccentricity of endosome, (J) concentration of Golgi apparatus, (K) compactness of Golgi apparatus distribution, and (L) eccentricity of Golgi apparatus distribution.

morphological quantification with high accuracy at the single-cell level, which is the key element for phenotyping and high-content screening [78].

3.5 Summary

Beyond predicting IF labels, I have further demonstrated quantitative cytometry analysis based on the multiplexed digital output from my DL models. Importantly, my analysis has shown that a multitude of single-cell profile metrics can be accurately extracted from the DL predictions. The digital multiplexing enabled us to simultaneously quantify several morphological features on multiple subcellular components across a large cell population. This capability drastically improves the technique's throughput for structural phenotyping in the application of imaging cytometry, such as high-content analysis/screening.

Chapter 4

Integrating Semi-supervised Digital Staining with Serial Sectioning Optical Coherence Tomography

4.1 Background

Optical coherence tomography (OCT) is an emerging label-free imaging technique that allows high-resolution 3D visualization and quantification of intrinsic optical properties of tissue, such as scattering coefficient, back-scattering coefficient [79, 13], and birefringence [80, 12]. Recently, OCT has shown great promise in brain imaging applications, such as visualizing single neurons [81], fiber tracts [14], and the laminar structure of the cerebral cortex in the human brain [82, 83]. While traditionally limited by light penetration, serial sectioning OCT (S-OCT) integrates OCT with a vibratome slicer to enable 3D imaging of cubic centimeters of tissue [84]. S-OCT permits simple and accurate 3D high-resolution reconstruction of large-scale brain anatomy, microstructures, and tractography [84, 85, 86] with minimal tissue distortion. This is achieved by the serial imaging protocol [87], which alternates between OCT imaging and slicing off the superficial tissue section, and enables preservation of the complex 3D structures of brain tissues without requiring sophisticated inter-slice registration algorithms. Despite its ability to routinely generate large-scale volumetric brain imaging data, S-OCT still requires domain expertise to identify and annotate neuronal features for further analysis [79, 14, 84, 11]. my goal is to augment S-OCT

with a digital staining (DS) technique that enables straightforward 3D pathology on large-scale human brain tissues.

In the past few years, deep learning methods have revolutionized the field of DS, which aims to transform label-free images into histological staining-like images using a computational model [19]. DS offers a fast and low-cost alternative to conventional physical staining (PS) methods. Several DS models have been developed that transform different pairs of input-output imaging modalities. However, most existing DS methods rely on supervised learning methods, which requires paired images of the tissue slice with and without staining for model training. To ensure accurate DS results, cross-modal registration between the image pairs with pixel-level accuracy is crucial [24, 25, 88, 36]. However, obtaining such image pairs is difficult and often involves sophisticated image registration procedures [19, 25]. To overcome this challenge, some recent studies have explored unsupervised image translation models for DS, which only need unpaired collections of images from the two modalities for model training [46, 47, 50, 48, 49]. The most popular unsupervised method is CycleGAN [44], which comprises two sets of generators and discriminators that enforce cycle consistency and content preservation for the image translation task. A recent improvement over CycleGAN is Contrastive Unpaired Translation (CUT) [45], which uses contrastive learning to achieve better structural and content preservation with only one set of generator and discriminator, and has demonstrated superior performance in DS tasks. However, these unsupervised models still lag behind supervised models in terms of accuracy.

4.2 Principles and Overview

Here I present a new semi-supervised learning framework for DS using a limited amount of weakly paired image data. As a proof-of-concept demonstration, I use

my DS model to translate S-OCT images to Gallyas silver staining. my DS model consists of two novel modules that address several challenges in my technique. my main model is based on the CUT framework to perform DS using unpaired training data. This module combines contrastive learning and adversarial learning to address the lack of paired imaging data since the physically stained images were obtained from unordered adjacent brain tissue sections to the OCT-imaged sections and were confounded by tissue damage and distortion during the staining process.

To improve the accuracy of the unsupervised model, I augment it with semi-supervision from two auxiliary tasks. Firstly, I devise a pseudo-supervised learning module by training the DS network on a pseudo-paired training dataset that is generated using my previously established biophysical model. my previous work has revealed a linear correlation between the OCT scattering coefficients (SC) and the optical density (OD) computed from the Gallyas silver stained image [11]. Based on this similarity prior, this module learns to translate the generated OD back to the Gallyas silver stain, acting as a proxy supervision for learning the translation from OCT-SC to Gallyas silver stain. This naturally pixel-aligned pseudo supervision augments the training data, enabling training the DS model effectively despite the limited data available to my task due to the scarcity of the human brain samples. Additionally, when combined with the adversarial learning component in the CUT backbone, the domain gap between the OCT-SC images and OD maps are effectively mitigated by the mechanism of domain-adversarial training [89]. Secondly, I develop an unsupervised cross-modality image registration module that aligns the adjacent Gallyas image with the OCT-SC image. This module enables the DS model to utilize the geometric similarity information provided by the adjacent slices, thereby guiding the image translation process. To train the registration network effectively, I introduce a novel two-stage, multiscale training strategy. It allows the network to

learn image registration at the “global” whole slide image (WSI) scale, while simultaneously learning image translation at the “local” image patch scale. Furthermore, this novel training strategy facilitates collaborative training between the DS model and the registration model, leading to more effective enforcement of high-quality DS results.

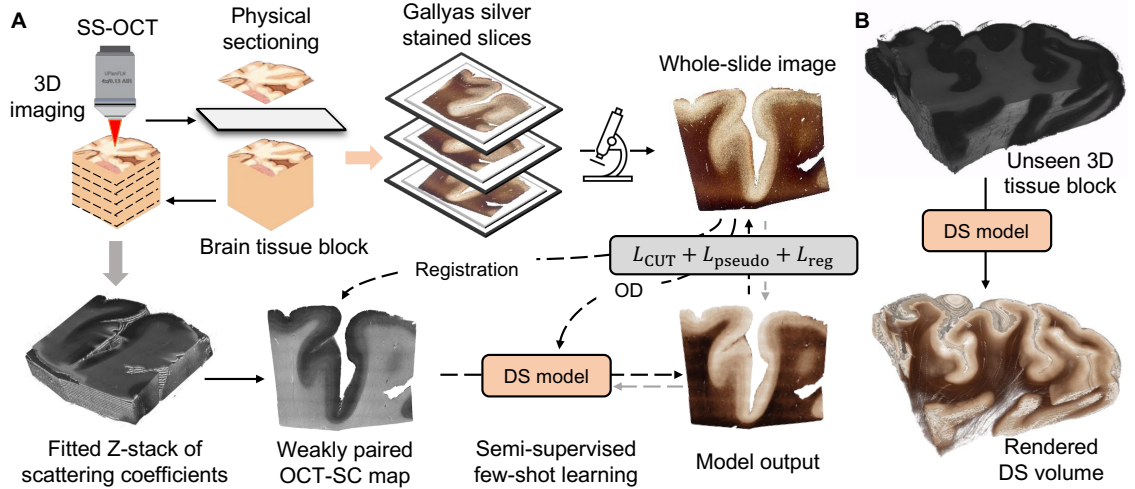


Figure 4-1: Overview of the proposed OCT DS technique. (A) Data acquisition and DS model. S-OCT alternates between 3D imaging and tissue sectioning to acquire a stack of block-face OCT images, which are then processed to compute the scattering coefficient (OCT-SC) map stack. Sectioned sample slices are physically stained and imaged. The DS neural network is trained from a few weakly-aligned pairs of OCT-SC and Gallyas silver-stained images. (B) After the DS model is trained, it can perform inference on completely new slices of OCT-SC images for volumetric DS.

I present my DS pipeline for data acquisition and deep learning model training in Fig. 4-1. I use S-OCT to obtain label-free volumetric data of human brain samples. I then process the OCT measurements to calculate the OCT scattering coefficient (SC) maps [79]. Next, I develop a deep learning DS model that transforms OCT-SC images into Gallyas silver stain images. I choose OCT-SC as the input for the DS model instead of the raw OCT measurements because SC measures the intrinsic optical properties of the tissue and eliminates the inhomogeneity in the raw scattering-

based OCT measurements by using a nonlinear model-fitting process [79]. Moreover, a biophysical model from my previous work showed that OCT-SC mainly depends on the contribution of myelin content and neuron distribution, which are captured by the OD of the Gallyas silver staining [11]. I hypothesize that the correlation between these two modalities can be leveraged to create a more accurate image-to-image mapping using a deep learning model. I also collect a few physically stained brain tissue slices for DS model training and evaluation. The deep learning model is trained on a few weakly-aligned pairs of OCT-SC and Gallyas silver stained whole slide images (WSI). The inference stage of the DS model is shown in Fig. 4.1B. After the model is trained, it can be applied on any OCT-SC maps to enable 3D neurohistology on cubic centimeters of brain tissue and visualize mesoscopic brain structures.

I present the OCT DS results on single-section tissues from various cerebral cortex samples and compare them with PS results from adjacent sections. I demonstrate that DS exploits the quantitative nature of OCT-SC and thus can produce consistent staining quality across different samples. Compared to PS, DS reveals comparable mesoscopic ($\sim 10 \mu\text{m}$) structures in different tissue regions without introducing staining variability across samples and experiments.

4.3 Methods

4.3.1 Serial-sectioning OCT platform and data processing

Serial-sectioning OCT

The S-OCT microscope was described previously [12]. I use a swept light source (AxsunTech) with 100kHz swept rate, a central wavelength of 1310 nm, and a spectral full width half maximum (FWHM) of 110 nm, yielding an axial resolution of 5.6 μm in brain tissue ($n=1.4$). I use a free space interferometer and quarter wave plate (QWP) to illuminate the sample with circularly polarized light, and use two balanced

detectors for measuring orthogonally polarized reflection light. A variable retarder placed in the sample arm is used to compensate for the system birefringence and to recover precise measurement of sample birefringence. To sample the spectrum in even-k space, I input the k-clock of the light source into a high-speed digitizer (ATS9350, AlazarTech), afterwards real-time FFT is carried out using a Graphic Processing Unit (RTX4000, NVIDIA), and the spatial-domain data is trimmed to only save the first 1 mm depth. The post-objective power was measured to be 3.7 mW, achieving a 95 dB SNR in both polarization channels. I use $1 \times 1 \text{ mm}^2$ FOV with $3 \text{ }\mu\text{m}$ lateral step size and 30% overlap between tiles. The sample is mounted on XYZ motorized stages which translate the sample to image the whole surface as well as between the vibratome and objective. After block-face imaging, a custom vibratome cuts off the $50 \text{ }\mu\text{m}$ slices with 0.3 mm/s cutting speed and 3000 rotations per minute (RPM) blade vibrating frequency.

Optical properties estimation with S-OCT

Tissue optical properties will be extracted by following a previously established procedure, to analyze the reflectance intensities of PSOCT using a nonlinear fitting method [79, 13]. To summarize, spatial parametrization is firstly applied to the confocal parameter across a 3D PSOCT image to constrain and reduce the degrees of freedom in the nonlinear coefficient fitting problem, resulting in improved confidence in the estimated optical properties of the sample. Afterwards, a nonlinear least-squares solver is used to estimate the back-scattering and scattering coefficients from the nonlinear reflectance-vs-depth over about $150 \text{ }\mu\text{m}$. All curve fitting is performed in MATLAB. After extracting the optical properties for each image tile, the tiles are stitched using the coordinates generated during the volumetric reconstruction with FIJI software [90].

4.3.2 Sample preparation and staining experiments

Sample preparation

15 samples were obtained from Boston University Alzheimer’s Disease Research Center brain bank. 5 of them were late stage Alzheimer’s disease (AD), 5 of them were late stage Chronic Traumatic Encephalopathy (CTE), the other 5 are age-matched normal controls. These samples were used for studying neurodegenerative changes in scattering and birefringence of brain tissue. These samples were fixed by immersion in 10% formalin for at least two months. The post-mortem interval did not exceed 24 h. Before imaging, the samples were washed in $1 \times$ phosphate buffered saline for a month to remove residual fixation agents and then embedded in 4.5% agarose for tissue support [91]. During embedding, the brain blocks were warmed to $65 \text{ }^\circ\text{C}$ to allow sufficient penetration of agarose into the deep sulcus. During imaging, the brain tissue block is mounted in a water bath filled with Deionized (DI) water. The DI water was changed every day to remove the debris from cutting that could degrade the OCT image quality. The collected tissue slices were stored in $1 \times$ PBS with antibacterial agent (sodium azide) at $4 \text{ }^\circ\text{C}$. I kept 1 slice per glass vial to keep the sequence of slices.

Gallyas silver staining and imaging

Total number of 35 brain slices from 15 samples (at least two slices per sample, slices were separated in depth by 1 mm to assure anatomical difference), with a thickness of $50 \text{ }\mu\text{m}$, were mounted onto gelatin-coated slides and subjected to Gallyas staining protocol, as described by Pistorio [92]. In light of the increased thickness of the samples, the impregnation and bleaching time were modified accordingly. Subsequently, the samples were captured in brightfield mode with a $10 \times$ objective (NA=0.4) and RGB camera using the VS-120 slide scanner for $75 \times 25 \text{ mm}$ slides. The exposure

time was 1.73 ms and the pixel size was $0.7 \mu\text{m}$ with $1 \times 1 \text{ mm}$ FOV. Wider samples were imaged using the BZ-X microscope under similar settings. The image can be opened in Olympus Olyvia and exported as tiff images for processing.

4.3.3 Image processing

Our image dataset consists of two types of images: PS images from the slide scanner and OCT-SC images computed from S-OCT. I first inspected all the PS images visually and excluded the ones that had low staining quality or artifacts in the training dataset. I selected 9 out of 35 PS WSIs for training my DS model. The PS WSIs had different sizes depending on the tissue sample, but they were around the median scale of 36 mm to 48 mm with the pixel size of $1.9 \mu\text{m}$. To generate the weakly-paired training dataset, I manually paired the PS images with the OCT-SC images that had the most similar appearance. Since the sectioning thickness ($50 \mu\text{m}$) of PS samples did not match the fitting thickness used for OCT-SC images ($150 \mu\text{m}$) and the depth information of PS samples was not recorded, I can only pair the PS with the closest adjacent OCT-SC image sample by qualitatively assessing the similarity of tissue features. I then downsampled the PS images using bicubic interpolation by a scale factor of 6.32 to match the $12 \mu\text{m}$ pixel size in OCT-SC images. I also cropped or padded the PS images to have the same image size as the corresponding OCT-SC images, which was around 3000×4000 pixels for each sample. I performed this procedure on all PS images when I compared them with the OCT-SC images side-by-side in my results.

The PS images undergo several preprocessing steps to minimize the effects of sample and staining variations before they are used for training. The preprocessing steps include background removal, intensity normalization and color transfer. The background removal eliminates the unwanted image artifacts in PS image and is done by interactive image segmenter in Matlab. The intensity normalization adjusts

the PS images to balance the varying illumination levels across different imaging experiments. The brightest pixel (I_r, I_g, I_b) is used to estimate the illuminant color and the image is scaled by the constant $\left(\frac{1}{I_r}, \frac{1}{I_g}, \frac{1}{I_b}\right)$ for each color channel, followed by a range normalization to map the overall image value range to $[0, 1]$. The color transfer uses Reinhard method [93] to standardize the staining color variations among experiment, sample and imaging conditions given a target PS image with a relatively ideal staining as reference.

The OCT-SC images obtained from the fitting algorithm show some artifacts mainly in the background areas and near the sharp boundaries of the vessel regions, because the algorithm assumes a constant SC value for the 150 μm imaging thickness [79]. To reduce the background noise and correct the over-smoothed values near the vessel edges, the OCT-SC images are further processed by several steps. First, the background is removed by applying a histogram-based thresholding method using the triangle algorithm [94], followed by a sequence of smoothing morphological operations such as erosion, small object removal and dilation. Next, the pixels with zero values in the masked image are identified as defective and are replaced by the local median. Then, the edges of the vessel regions are detected using a difference-of-Gaussian (DoG) filter and thresholding. Finally, the outlier regions with small values compared to the local maximum are segmented and combined with the edge mask. The combined mask is smoothed by similar morphological filters, and the values in the mask are replaced by the local maximum. The preprocessing pipeline is implemented in Python using the basic filters and morphological operators from scikit-image package [94].

For creating a 3D visualization of the DS images that show the volumetric digital staining results, I change the white-color background of the DS images to black, so that only the sample region is visible. This is done by converting the DS color im-

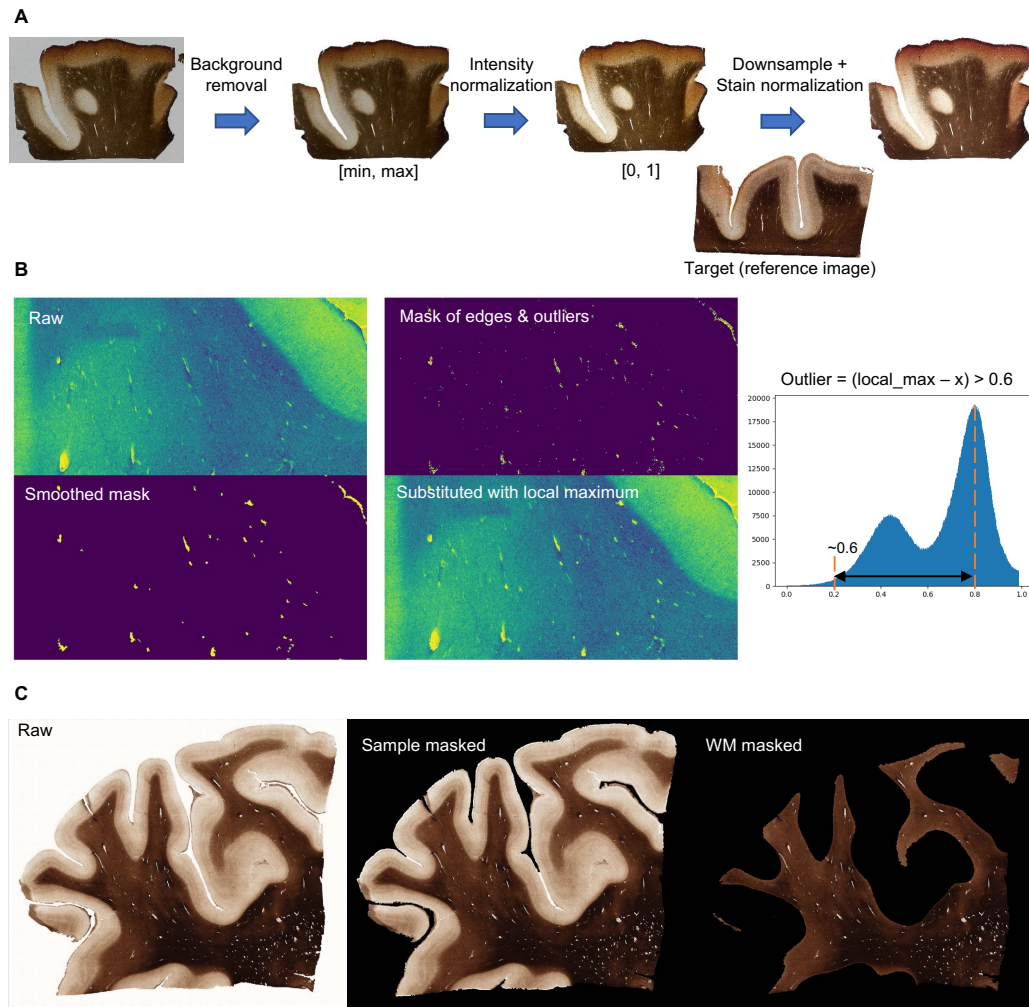


Figure 4.2: Image processing pipeline. (A) Preprocessing of PS images includes background removal, intensity normalization, resolution matching (downsampling) and color transfer. (B) Preprocessing of OCT-SC images include edges & outliers masking, mask smoothing and maximum filtering. The outlier mask is determined by comparing pixel values to the local maximum shown on the right panel. (C) Post-processing of DS images. The second and third column shows results with sample masked and white matter (WM) region masked.

ages to grayscale and applying a triangle method threshold to select the foreground pixels. Then, a morphology smoothing operation is performed to remove any noise or artifacts. To extract the white matter (WM) masks from the DS grayscale images for highlighting the WM regions in the sample, I use a histogram thresholding method

based on the minimum method [94] and apply another morphology smoothing operation. The pixels that are not part of the WM masks are set to zero, and the resulting images are stacked in a volume for 3D visualization. The 3D viewer in ImageJ [90] is used to display the volume.

Fig. 4.2 serves to further elucidate the image processing pipeline employed for OCT-SC, DS, and PS images, where each subfigure provides a visual representation of the specific preprocessing steps applied to the respective image modality. In Fig. 4.2A, I present the preprocessing steps implemented for PS images. This involves background removal, intensity normalization, and color transfer techniques to enhance the visual quality and ensure consistency across images. Fig. 4.2B focuses on the preprocessing steps tailored for OCT-SC images. The aim here is to mitigate the background interference and correct the over-smooth values typically observed around vessel region edges. These processing steps contribute to improving the overall clarity and accuracy of the OCT-SC images. Finally, Fig. 4.2C showcases the post-processing steps carried out on DS images. This includes sample masking and the application of white matter (WM) masking. By employing these techniques, I am able to isolate and highlight specific areas of interest within the DS images, facilitating a more detailed analysis.

To generate the training image dataset, I used PyTorch to create a parallel processing module that can split the WSIs of different image sizes into smaller patches during training on the fly. This allows us to dynamically update the intermediate image tensors that can be input to different parts of deep learning models to train at different image scales. The WSIs dataset with different sizes can then be directly handled by a custom data loader for standard-size tensor operation. I first pad the WSI to the size of multiple integers of patch size, and then use the tensor unfolding method in PyTorch to cut the image tensor using a sliding window into smaller ten-

sors stacked in the batch dimension. The inverse stitching operation is done similarly using the tensor folding methods.

4.3.4 Semi-supervised deep learning framework

Problem formulation and framework design

I formulate the DS task as a weakly-paired image translation problem because I do not have access to pixel-aligned image pairs of OCT-SC and PS images. To achieve better performance than fully unsupervised methods, I exploit the side information provided by the structural and content similarity between the adjacent sections in the imaging data, as well as a biophysical model for linking OCT-SC and the contrast in Gallyas silver stain in a semi-supervised deep learning framework.

The training framework of my DS network consists of several innovative learning components, as shown in Fig. 4-3. Based on the CUT framework as the backbone [45], the DS model uses a mix of adversarial loss and contrastive loss in the unpaired image setting, as shown in Fig. 4-3A. The adversarial learning measures the perceptual similarity of the generated DS images and the PS images. It tries to reduce the gap between the high-dimensional distributions of the DS and PS images such that the generated DS images are perceptually indistinguishable from the PS images. The contrastive loss uses self-supervised patch-wise learning to ensure structural consistency between the OCT-SC and DS images. It maximizes the mutual information and provides self-guidance for content preservation. The combination of contrastive loss and adversarial loss enables high-quality DS images that preserve the content and structures of the OCT-SC images.

To improve upon the unsupervised CUT framework, I propose a semi-supervised learning method. my method leverages augmented pseudo pairs generated by a biophysical model and registered cross-modality image pairs that are dynamically adjusted by a learnable registration network. The intuition is that using additional

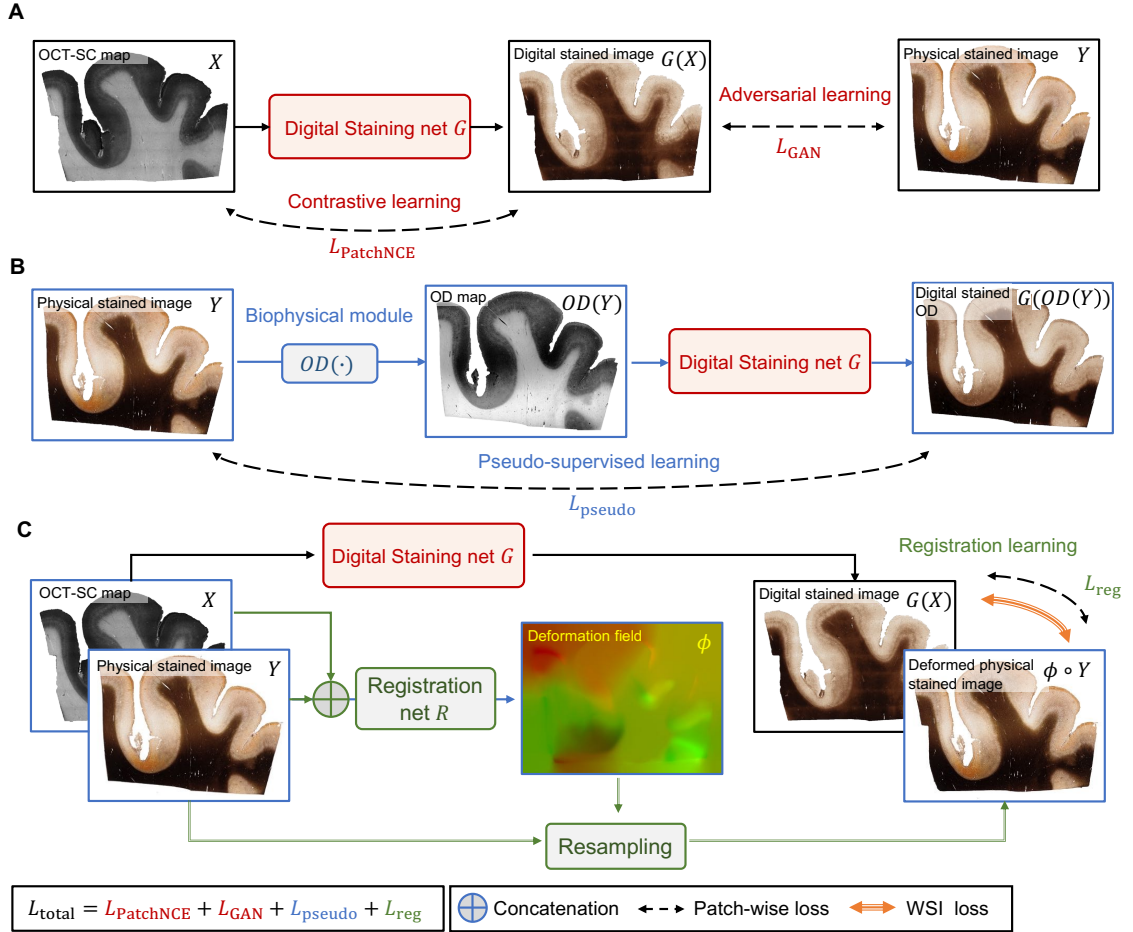


Figure 4-3: The training framework of my DS neural network model. (A) The backbone of the DS network G is built on the CUT framework, which combines contrastive learning and adversarial learning. The input is a 2D OCT-SC map X and the output is a digitally stained image $G(X)$ that is compared with a PS image Y from an adjacent slice. (B) Auxiliary pseudo-supervised learning task. The biophysical module computes the optical density $OD(Y)$ of the PS image Y , which is fed as an input to G . The digitally stained OD image $G(OD(Y))$ is compared with the original PS image Y during training. (C) Auxiliary unsupervised cross-modality image registration task. I alternate between optimizing G and a registration network R under different image scales. I fix R while updating G , which provides more informative supervision for R in the next iteration. I use patch-wise losses for training G , and WSI losses for training R .

auxiliary supervision enhances the learnability, efficiency and accuracy of the model compared to unsupervised learning. Crucially, my semi-supervised method does not

require any exact paired PS and OCT-SC images during training.

In Fig. 4.3B, I introduce the pseudo-supervised learning auxiliary task to enhance the unpaired image translation for DS of OCT-SC images. I first compute the OD maps from the PS images and then utilize the OD - PS image pairs to train the DS model in a pseudo-supervised manner. This approach proves effective because the OD image exhibits similar image contrast and feature distribution as the OCT-SC across various cortical regions. Additionally, the OCT-SC demonstrates an approximate linear relationship with the OD of the Gallyas silver stain [11]. Furthermore, since the OD map is naturally pixel-aligned with the PS image, it facilitates supervised learning and provides additional semi-supervision and alignment constraints for the main DS model. However, the inherent disparities in image features and intensity value distributions between the OD map and the OCT-SC image result in a domain gap, which limits the accuracy of the trained DS model when relying solely on this auxiliary task. My insight is that when this task is combined with the adversarial learning component in the CUT backbone, it enables domain adaptation similar to the domain-adversarial training framework [89]. The performance on the OCT-SC image is ensured by penalizing the perceptual differences between the DS images generated from the OCT-SC image and the OD map using the adversarial loss. By leveraging both the pseudo-supervised learning and adversarial learning components, I effectively bridge the domain gap and improve the accuracy of the DS model for OCT-SC image translation.

In Fig. 4.3C, I illustrate the second auxiliary task for aligning the PS image, the OCT-SC image, and the DS image using a registration network. This registration module undergoes two training stages: pre-training and fine-tuning. During the pre-training stage, the registration module operates on the WSI scale. It predicts a deformation field that indicates the pixel-wise displacement vectors required for non-

rigid transformation. To facilitate cross-modal self-supervised registration, I utilize the OD map as a surrogate for the OCT-SC image and learn a deformation field between the OD map and the input OCT-SC image. This result is used as an initial estimate for the deformation between the PS image and the matching OCT-SC image. By leveraging my biophysical model, I bootstrap the challenging self-supervised cross-modality image registration problem in this pre-training stage. The subsequent fine-tuning of the registration model aims to provide pixel-wise weak-supervision for the DS model. In this stage, I employ an alternate training approach that involves collaborative learning between the DS model and the registration model. When the DS model is fixed, the registration model is trained at the WSI scale to address global geometry correction. When the registration model is fixed, the DS model is trained at the image patch scale to provide sufficient samples for local translation learning. This unsupervised cross-modality image registration module enables the DS model to learn improved local color tone mapping from unaligned imaging modalities without the need for explicit supervision.

Overall, my DS framework augments unpaired image translation with pseudo supervised learning and unsupervised cross-modality image registration. The total loss function used for training is the weighted sum of the four objectives derived from the main image translation task and two auxiliary tasks.

Loss functions and implementation details

The proposed framework combines generative adversarial learning, contrastive learning, pseudo supervised learning based on self-generated image pairs based on a biophysical model, and unsupervised cross-modality image registration.

I denote the OCT-SC images as X and the PS images as Y . The main framework consists of a DS network G and a registration network R . The DS network G transforms grayscale OCT-SC images X into color images that resemble the color and

contrast of PS images Y . The registration network R takes pairs of unaligned images X and Y as input and outputs a deformation field $\phi = R(X, Y)$ that can be applied to resample and register Y to X . I use an auxiliary discriminator network D to enforce perceptual similarity between the output DS and reference PS images by adversarial learning. I also use a fully connected network f to ensure structural consistency between the input OCT-SC and output DS images by contrastive learning.

My framework operates on two different image scales: WSI scale (denoted by upper case letters) and image patch scale (denoted by lower case letters). R is trained on WSIs, which have a size of about 3000 to 4000 pixels. G , D , f are trained on image patches, which have a size of 512×512 pixels. I design an efficient image processing module to either split (X, Y) into patches (x, y) or stitch patches back to WSIs, as detailed in the Image Processing section. The CUT framework is used to jointly train the networks G , D , and f during the training phase. Additionally, G undergoes a pseudo-supervised training scheme and an alternating training process with R , which are explained below.

In Fig. 4.4A, I demonstrate the semi-supervised training of the DS generator G , which combines contrastive learning, adversarial learning, pseudo-supervised learning, and learnable registration. The OCT-SC map X and the corresponding PS image Y from an adjacent slice are utilized during training, but no exact supervised pairs of X and Y are explicitly provided to G . The training process involves alternate optimization of G and the learnable registration component R at different image scales. R is held fixed while optimizing G , and the updated G provides informative supervision for the subsequent iteration of R . Patch-wise losses are formulated and combined into a total loss during the training of G .

In Fig. 4.4B, I illustrate the training of unsupervised cross-modality image registration. WSI losses are formulated between different pairs of image modalities at two

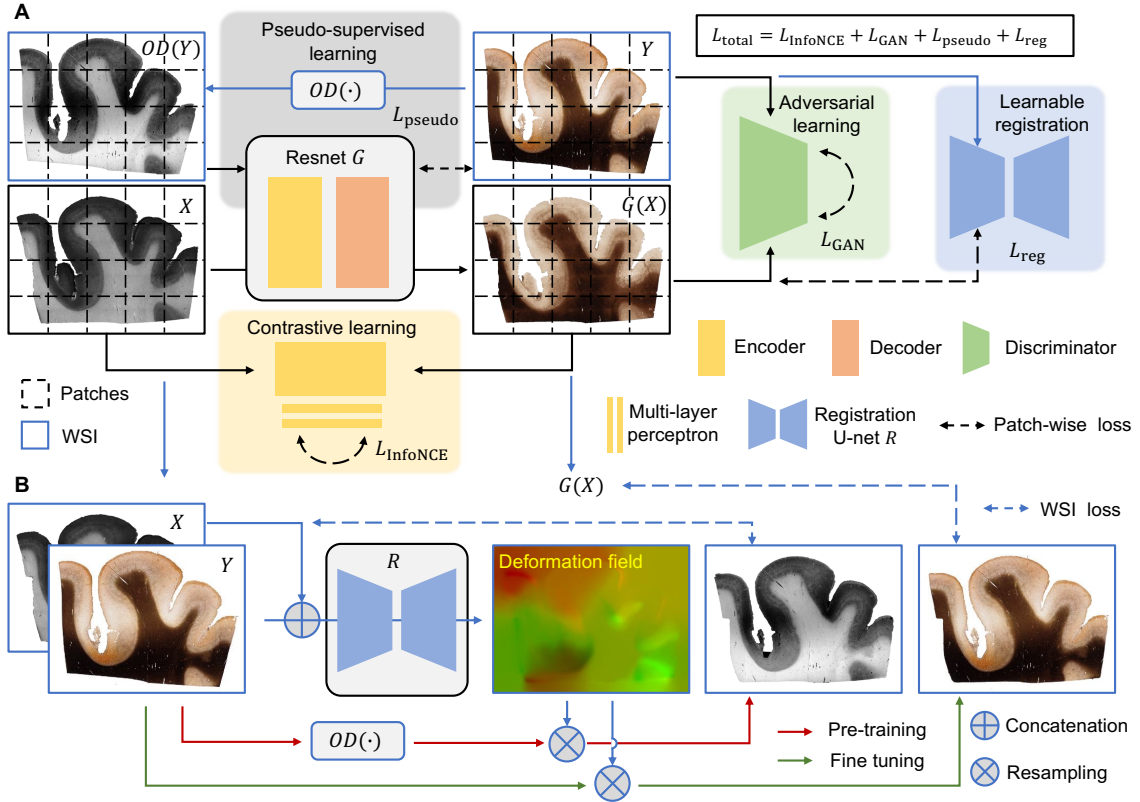


Figure 4.4: Detailed semi-supervised training framework of my DS model. (A) Semi-supervised training of DS generator combining contrastive learning, adversarial learning, pseudo-supervised learning and learnable registration. Patch-wise losses are formulated and combined in the total loss. Blue arrows and blue boxes indicate images are processed at the WSI scale. Black arrows, dotted black double-arrows and dotted black boxes indicate images are in patch-wise scale and are processed patch-wise. (B) Unsupervised cross-modality image registration. WSI losses are formulated between different image-modality pairs at two training stages. R is trained stand-alone at the pre-training stage, where similarity prior is used. At the fine-tuning stage, R is trained collaboratively with G using alternate optimization. Red arrows indicate the forwards pass during pre-training, and green arrows indicate the forward pass during fine-tuning. All images are processed as WSI scale in (B).

distinct training stages. At the pre-training stage, R is trained independently across the entire WSI dataset, utilizing a similarity prior between X and $OD(Y)$. During this stage, the focus is on establishing a robust registration model. At the fine-tuning stage, R and G are alternately trained using selected high-quality WSIs. This stage

enables the refinement of the registration model by leveraging the information provided by G .

The adversarial learning module maximize the perceptual similarity between the DS output $G(x)$ and the PS image y using an auxiliary discriminator D . D is trained to distinguish between the desired modality y and the generated images $G(x)$. The least-squares GAN loss is used to measure how well D outputs match the binary labels for both y and $G(x)$. The training procedure alternates between two steps: fixing G and updating D with $L_{\text{GAN}}(D)$, and fixing D and updating G with $L_{\text{GAN}}(G)$:

$$L_{\text{GAN}}(D) = \mathbb{E}_y (D(y) - 1)^2 + \mathbb{E}_x (D(G(x)))^2 \quad (4.1)$$

$$L_{\text{GAN}}(G) = \mathbb{E}_x (D(G(x)) - 1)^2 \quad (4.2)$$

The contrastive learning module ensures that the semantic structure of x is preserved when it is translated to $G(x)$. I implement G with a ResNet model and treats the first half of the ResNet layers as the encoder and the remaining layers as the decoder. The encoder G_{enc} transforms images from both domains into a common latent space, and the decoder G_{enc} generates translated images from latent vectors. To formulate the multi-layer patch-wise contrastive loss, I adopt the approach in [45] to sample the encoded feature maps from both x and $G(x)$. Each layer and spatial location in the feature map stack corresponds to a patch of the input image that covers the corresponding receptive field. I extract multiple layers of the encoded feature maps, randomly sample the spatial locations and apply a fully connected network f to obtain a stack of latent features $z_{s,l} = f(G_{\text{enc}}^{s,l}(x))$, where s is the spatial index within $[1, S]$ and l is the selected layer within $[1, L]$. I do the same processing on image $G(X)$: $\hat{z}_{s,l} = f(G_{\text{enc}}^{s,l}(G(x)))$. Then I compute a PatchNCE loss using a cross-entropy contrastive loss formulation:

$$L_{PatchNCE}(G, f, x) = \mathbb{E}_x \sum_{l=1}^L \sum_{s=1}^S -\log \left(\frac{\exp(z_{s,l} \cdot \hat{z}_{s,l})}{\sum_{t=1}^S \exp(z_{s,l} \cdot \hat{z}_{t,l})} \right) \quad (4.3)$$

This loss function encourages the latent representations of image patches from x and $G(x)$ that belong to the same spatial location to have similar content to be close in the feature space, while pushing away the representations of patches that are uncorrelated or have different content. By this internal negative sampling scheme in the feature space, the model learns to contrast positive and negative pairs of patches based on their content similarity, which maximizes the mutual information between the input image x and the output image $G(x)$. This provides a self-supervised signal for preserving the content of the image during the transformation.

The pseudo-supervised training is formulated as a pixel-wise loss function that minimizes the discrepancy between the translated optical density (OD) and the physical Gallyas-silver stain (PS) images. To enable this training, I first compute the OD of image Y by

$$OD(Y) = -\frac{1}{3} \sum_{c=R,G,B} \log_{10} Y_c \quad (4.4)$$

However, there is a mismatch between the intensity values from X and $OD(Y)$. The domain difference between these input modalities hinders the model’s generalization on X if the model is directly trained on pairs of $OD(Y)$ and Y . Therefore, I apply histogram equalization on WSIs of $OD(Y)$ and X before passing them to G . Then I extract patches $OD(y)$ and y from the processed WSIs and apply the auxiliary pseudo-supervised loss as follows:

$$L_{Pseudo}(G) = \mathbb{E}_y \|G(OD(y)) - y\|_1 \quad (4.5)$$

The cross-modality image registration module is trained in two stages. In the first stage, I pre-train the registration network R on WSIs of X , Y and $OD(Y)$. The

registration network R takes weakly-paired X and Y as input and predicts a deformation field $\phi = R(X, Y)$ that indicates the pixel-wise displacement vectors needed to perform the non-rigid transformation. To formulate a cross-modal self-supervised registration loss L_{reg}^I , I use $OD(Y)$ as a surrogate of Y and exploit the correlation between OD map of the PS image $OD(Y)$ and the input OCT-SC image X . By minimizing the difference between the registered $OD(Y)$ and X , I indirectly learn the deformation between Y and X . This training is enabled by a differentiable resampling layer that performs image warping denoted by \circ . I also add a total variation (TV) regularization term to encourage the smoothness of the learned deformation field. The registration loss during this pre-training stage is computed at the WSI scale as follows:

$$L_{\text{reg}}^I(R) = \mathbb{E}_{X,Y} \|X - \phi \circ OD(Y)\|_1 + \|\phi\|_{TV} \quad (4.6)$$

where $\|\phi\|_{TV}$ is the total variation norm calculated as:

$$\|\phi\|_{TV} = \sum_{i,j} \sqrt{|\phi_{i+1,j} - \phi_{i,j}|^2 + |\phi_{i,j+1} - \phi_{i,j}|^2} \quad (4.7)$$

In the second fine-tuning stage, I train R and G in an alternating and collaborative manner. The purpose of fine-tuning R is to provide pixel-wise weak-supervision between the registered Y and the DS image $G(X)$, which in turn helps to fine-tune the DS network G . Using the coarsely-trained G , I can produce $G(X)$ that has the same image modality as the PS image Y and use a pixel-wise loss function to perform training. I implement the following scheme for alternating training. When I fix G , I train R by comparing the registered PS image Y and the DS image $G(X)$ at the WSI scale using the loss function:

$$L_{\text{reg}}^{II}(R) = \mathbb{E}_{X,Y} \|G(X) - \phi \circ Y\|_1 + \|\phi\|_{TV} \quad (4.8)$$

When I fix R , I crop the intermediate registered WSI Y into patches $\phi_y \circ y$ and train

G at the WSI scale by comparing the registered PS image patch and the DS image patch $G(x)$ using the loss function

$$L_{\text{reg}}^I(G) = \mathbb{E}_{x,y} \|G(x) - \phi_y \circ y\|_1$$

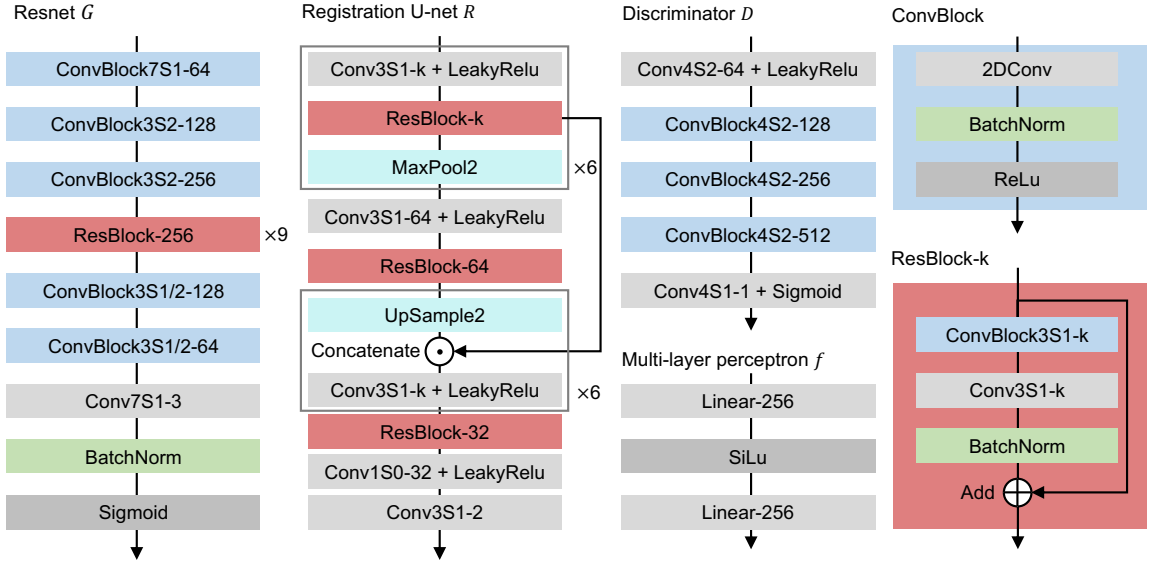


Figure 4-5: Model architecture details. The building blocks are convolutional blocks (ConvBlock) and residual blocks (ResBlock). The naming rules are as follows: Conv(i)S(j)-(k) represents an ii 2D convolution with k filters and stride j, where fractional stride size indicates transpose convolution; ConvBlock(i)S(j)-(k) represents a block consisting of an ii 2D convolution with k filters and stride j, followed by a batch normalization (BatchNorm) layer and a ReLu activation. ResBlock-(k) represents a ConvBlock3S1-k followed by a Conv3S1-k and BatchNorm; Linear-k is a fully connected layer with k hidden neurons. All ReLus in ConvBlocks of discriminator D are leaky ReLu. I use all leaky ReLus with a slope of 0.2.

The network structures of each component model are depicted in Fig. 4-5, providing a comprehensive overview of the architecture employed in each model. The generator G follows the ResNet architecture with 9 ResBlocks. The registration net R is a ResU-net with 6 downsampling blocks. The discriminator follows the structure of 70×70 PatchGAN. The Multi-layer Perceptron (MLP) used in contrastive loss is a simple two-layer fully connected network with 256 hidden neurons in each layer.

4.3.5 Training details of all baseline models for comparisons

In order to ensure a fair comparison, all deep learning models in this study utilize the same ResNet structure as the backbone for the DS generator G . Additionally, uniformity is maintained among the discriminators, which follow the PatchGAN structure, and the multi-layer perceptron (MLP), which adopts a consistent two-hidden-layer structure. To identify the optimal model, I conducted empirical optimization of hyperparameters and determined the appropriate number of epochs for early stopping (out of a total of 500 epochs). My optimization primarily focused on adjusting the loss weights and learning rate through a systematic grid search approach. In terms of the learning rate scheduler, I compared different decay strategies, including linear decay, exponential decay, and cyclic learning rate. After careful evaluation, I found that a linear decay schedule following the first 50 epochs yielded the best results across all models.

As for the first baseline model, I follow the model architecture of the original CycleGAN [44] and utilize a combination of adversarial loss and cycle consistency loss to train two pairs of generators and discriminators (G, D_y) and (F, D_x) . To ensure fair comparison, the adversarial loss is the least-squares style GAN loss that is also used in all other models as shown previously:

$$L_{\text{GAN}}(D_y) = \mathbb{E}_y (D_y(y) - 1)^2 + \mathbb{E}_x (D_y(G(x)))^2 \quad (4.9)$$

$$L_{\text{GAN}}(G) = \mathbb{E}_x (D_y(G(x)) - 1)^2 \quad (4.10)$$

The cycle consistency loss is formulated as:

$$L_{\text{cyc}}(G) = \mathbb{E}_x \|F(G(x)) - x\|_1 \quad (4.11)$$

$$L_{\text{cyc}}(F) = \mathbb{E}_y \|G(F(y)) - y\|_1 \quad (4.12)$$

yielding a final full objective of

$$L_{\text{CycleGAN}}(G, F, D_x, D_y) = L_{\text{GAN}}(G, D_y) + L_{\text{GAN}}(F, D_x) + \lambda_a L_{\text{cyc}}(G) + \lambda_b L_{\text{cyc}}(F) \quad (4.13)$$

with two hyperparameters λ_a and λ_b . In my training experiments, I fine-tuned the CycleGAN model and selected the best parameter setting as $\lambda_a = \lambda_b = 10$. The optimized learning rate is 0.0002. I ran 500 epochs of training and used quantitative metrics to monitor the model performance during training. The best model is selected at epoch 465.

For the CUT and FastCUT models, the full objective is a combination of least-squares GAN loss and PatchNCE loss:

$$L_{\text{CUT}}(G, D, f) = L_{\text{GAN}}(G, D) + \lambda_x L_{\text{PatchNCE}}(G, f, x) + \lambda_y L_{\text{PatchNCE}}(G, f, y) \quad (4.14)$$

As for the CUT model, the PatchNCE loss here is computed both on the input images x and on the output images y with two hyperparameters λ_x and λ_y . I fine-tuned and selected the best parameters as $\lambda_x = 20, \lambda_y = 10$. The optimized learning rate is 0.0003. Out of 500 epochs of training, the best model is selected at epoch 190. For the FastCUT model, the PatchNCE loss is only computed on the input images x with one hyperparameter λ_x . I fine-tuned and selected the best parameter as $\lambda_x = 20$. The optimized learning rate is 0.0002. The best model is reached at epoch 280.

In my proposed model, the final objective for training G is a combination of adversarial loss, PatchNCE loss, pseudo-supervised loss and registration loss:

$$L_{\text{Ours}}(G, D, f) = L_{\text{GAN}}(G, D) + \lambda L_{\text{PatchNCE}}(G, f, x) + \alpha L_{\text{Pseudo}}(G) + \beta L_{\text{reg}}^H(G) \quad (4.15)$$

I fine-tuned and selected the best parameters as $\lambda = 20, \alpha = 10, \beta = 1$. The optimized learning rate is 0.0002. The best model is selected at epoch 260. In addition, when

training the model R for registration, I used the following loss function at the pre-training stage:

$$L_{\text{reg}}^I(R) = \mathbb{E}_{X,Y} (\gamma \|X - \phi \circ OD(Y)\|_1 + \eta \|\phi\|_{TV}) \quad (4.16)$$

At the fine-tuning stage, the loss function is:

$$L_{\text{reg}}^{II}(R) = \mathbb{E}_{X,Y} (\gamma \|G(X) - \phi \circ Y\|_1 + \eta \|\phi\|_{TV}) \quad (4.17)$$

The optimized loss weights for both stages are $\gamma = 100, \eta = 600$ with the learning rate 0.0002.

4.3.6 Customized quantitative metrics

I first perform pairwise quantitative evaluations of the DS results using traditional image metrics such as mean squared error (MSE), Pearson correlation coefficient (PCC), and color difference (CD). Pairwise comparisons are carried out between whole-slide DS images and pre-registered PS images. Here, for simplicity, I denote the DS modality as X_i^c and PS modality as Y_i^c , where $c \in \{R, G, B\}$ represents different color channels in standard RGB (sRGB) space, and $i \in [1, N]$ indicates pixel indices.

The MSE is calculated as:

$$\text{MSE}(X, Y) = \frac{1}{3N} \sum_{c \in \{R, G, B\}} \sum_{i=1}^N (X_i^c - Y_i^c)^2 \quad (4.18)$$

The PCC is defined as:

$$\text{PCC}(X, Y) = \frac{\sum_{i,c} (X_i^c - \bar{X})(Y_i^c - \bar{Y})}{\sqrt{\sum_{i,c} (X_i^c - \bar{X})^2 \sum_{i,c} (Y_i^c - \bar{Y})^2}} \quad (4.19)$$

To quantify the color tone similarity in a perceptually uniform color space, I first convert X^c, Y^c from sRGB to CIELAB color space, where $c \in \{L, a, b\}$. The CD is

then defined as:

$$\text{CD}(X, Y) = \frac{1}{N} \sum_{i=1}^N \sqrt{(X_i^L - Y_i^L)^2 + (X_i^a - Y_i^a)^2 + (X_i^b - Y_i^b)^2} \quad (4.20)$$

I customize a metric to quantify the similarity of visible cortical layers in DS and PS images, specifically layer IV/V/VI. The layer mask extraction procedure can be found in Section 10. I denote the layer mask (binary image) of DS and PS modality as LX and LY , respectively. The intersection over union score of extracted cortical layer masks (IOU-layers) is defined as:

$$\text{IOU-layers}(X, Y) = \frac{|LX \cap LY|}{|LX \cup LY|} \quad (4.21)$$

where $||$ means the sum of absolute values, \cap and \cup indicates the intersection and union operation respectively.

I further customized unpaired metrics between whole-slide DS images and unregistered PS images. The myelin fiber (MF) segments masks MF_X and MF_Y are extracted. I compute the length and diameter of individual segments and then calculate the aggregated histograms of the two features as the approximate probability density distributions. The Jensen-Shannon divergence (JS) of MF length distributions p_L and q_L (from DS and PS respectively) is defined as:

$$\text{JS-MF length}(X, Y) = \frac{1}{2} \mathbb{D}_{KL} \left(p_L \parallel \frac{1}{2}(p_L + q_L) \right) + \frac{1}{2} \mathbb{D}_{KL} \left(q_L \parallel \frac{1}{2}(p_L + q_L) \right) \quad (4.22)$$

where $\mathbb{D}_{KL}(p||q) = \sum p(x) \log \frac{p(x)}{q(x)}$ is the asymmetric KL-divergence. Similarly the JS of MF diameter distributions p_D and q_D is defined by:

$$\text{JS-MF diameter}(X, Y) = \frac{1}{2} \mathbb{D}_{KL} \left(p_D \parallel \frac{1}{2}(p_D + q_D) \right) + \frac{1}{2} \mathbb{D}_{KL} \left(q_D \parallel \frac{1}{2}(p_D + q_D) \right) \quad (4.23)$$

The MF-segments number is counted as the number of individual segments ex-

tracted, where each segment is determined as one connected component (CC) with 2-connectivity.

$$\text{MF-segments number}(X) = |CC(MFX)| \quad (4.24)$$

The MF-area percentage is calculated by dividing the area of MF mask MFX by the area of sample mask SX :

$$\text{MF-area percentage}(X) = \frac{|MFX|}{|SX|} \quad (4.25)$$

4.3.7 Image analysis

In Fig. 4-6, I present the image analysis pipeline employed for myelin fiber extraction and quantification. The figure provides a step-by-step illustration of the process involved. Fig. 4-6A focuses on the segmentation of cortical layers IV/V/VI. Initially, the DS and PS images are converted to grayscale, and the initial sample mask is generated by applying thresholding using the triangle's method. The sample mask is then refined through the removal of small objects and small holes using a structure element of size 256 and 512, respectively. Erosion is subsequently performed with a square kernel of size 75. To enhance the efficiency of this operator with large kernel size, as an alternative approach, I employ an Euclidean distance transform on the binary sample mask image and apply thresholding with a threshold value of 75.

The next step involves the segmentation of a GM mask to exclude the white matter WM region. This is achieved through the minimum thresholding method, followed by the filling of white holes and dilation using a square kernel of size 8. To further refine the segmentation, a brown color mask is extracted to eliminate any remaining regions that are not stained brown. A squared error map is computed by subtracting the mean value of the WM region from the original image and squaring the residual. Thresholding is then performed on this map using the mean method, followed by

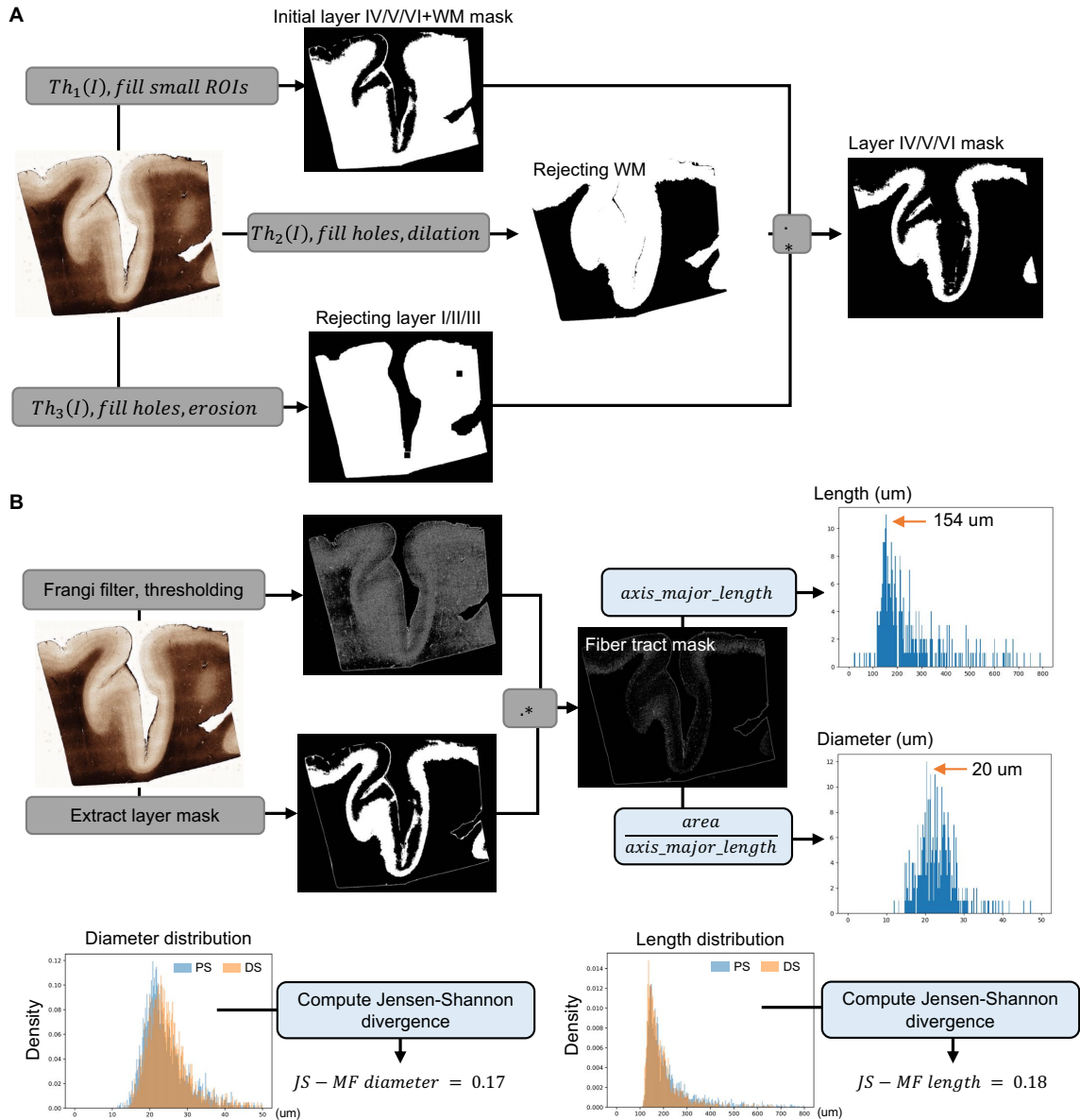


Figure 4-6: Image analysis for fiber segmentation and quantification. (A) The mask of cortical layer IV/V/VI is segmented by combining an initial layer + white matter (WM) mask, a gray matter (GM) mask and brown color mask. The layer masks are directly used for quantitative evaluation. (B) Frangi filter and layer mask are combined to segment fiber tracts within layer IV/V/VI. Myelin fiber features such as lengths and diameters are extracted to compute histograms. Further quantitative metrics such as Jensen-Shannon divergence are computed from two histograms of extracted features.

the removal of small holes and erosion. Finally, the three masks (sample mask, gray matter mask, and brown color mask) are combined through pointwise multiplication, resulting in an estimation of the mask representing the layer IV/V/VI regions. These layer masks are utilized for quantitative evaluation.

In Fig. 4-6B, the Frangi filter (followed by a thresholding operation) is applied to extract tubeness structures, which, in my case, predominantly correspond to myelin fibers. The Frangi filtered mask and the extracted layer mask are combined to segment the fiber tract mask within the layer IV/V/VI region. Subsequently, this mask is utilized to estimate the length and diameter of each individual fiber segment using the scikit-image property package. The length and diameter histograms are compared between the DS and PS images, and further analysis such as the computation of Jensen-Shannon divergence can be performed to assess the differences between the two distributions. Importantly, this analysis is carried out on population-level features, not necessarily on registered images, enabling unpaired quantitative evaluation.

4.4 Results

4.4.1 Digital staining by semi-supervised learning using weakly-paired images

I formulate the DS task as a weakly-paired image translation problem because I do not have access to pixel-aligned image pairs of OCT-SC and PS images. To achieve better performance than fully unsupervised methods, I exploit the side information provided by the structural and content similarity between the adjacent sections in the imaging data, as well as a biophysical model for linking OCT-SC and the contrast in Gallyas silver stain in a semi-supervised deep learning framework. My DS framework augments unpaired image translation with pseudo supervised learning and unsupervised cross-modality image registration. Overall, my method achieves superior performance

over other baseline methods, including CycleGAN, CUT and FastCUT in terms of DS quality and accuracy.

Qualitative comparisons against previous baseline models

I evaluate my semi-supervised approach against previous baseline methods. Fig. 4-7 shows whole slide images (WSIs) of OCT-SC, DS with four training methods and PS images from adjacent cortex sections. The baseline methods are CycleGAN, CUT and FastCUT models. my model outperforms them in resolution and color contrast similarity to PS images, as seen in the whole field-of-view (FOV) and zoom-in regions of interest (ROIs).

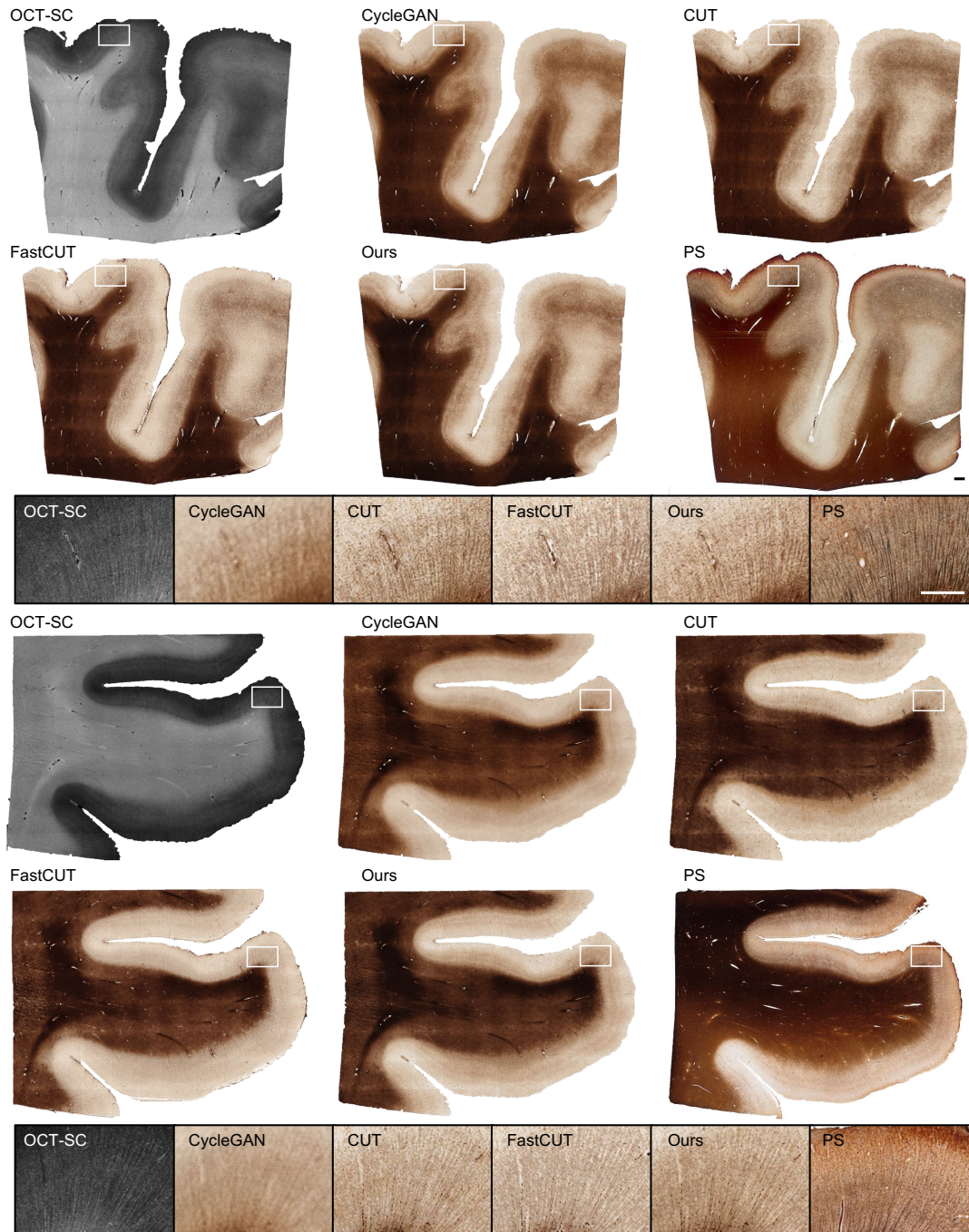


Figure 4-7: Qualitative comparisons against previous baseline models. Results from four methods are visualized: CycleGAN, CUT, FastCUT, and Ours. PS images from adjacent cortex tissue sections are shown as references. Zoom-in regions are marked by the white rectangular boxes in WSIs. Scale bars are 1 mm.

Quantitative comparisons to baseline models

I present the quantitative comparison results between my approach and other baseline methods in Fig. 4-7. To address the challenges posed by the weakly paired dataset, I have devised custom quantitative metrics for both pair-wise and unpaired evaluations. These metrics are visualized using box plots, which indicate the median and quantile values, employing five distinct colors to represent four different models and the PS images.

In Fig. 4-8A, I focus on pairwise metrics comparing registered PS images and DS images within the well-stained sample group to ensure fair comparisons. Lower values are desired for metrics such as mean squared error (MSE) and color difference (CD), while higher values are preferred for metrics like Pearson correlation coefficient (PCC) and Intersection over union of layers (IOU-layers). Among the four pairwise metrics, my method consistently outperforms three baseline models, exhibiting better scores (with CD comparable to FastCUT) and lower variations.

Moving on to Fig. 4-8B, I examine unpaired metrics that evaluate either the per-image distribution or the distribution of the entire dataset. These metrics are applied to samples across the entire sample group, with PS images remaining unregistered. I extract myelin fiber segments from the DS and (adjacent) PS images, computing histograms of different properties for comparison, such as fiber segment lengths (MF length) and fiber segment diameter (MF diameter). The divergence between the histograms of these properties is quantified using Jensen-Shannon divergence (JS). Lower values are desired for both JS-MF length and JS-MF diameter. my proposed model outperforms CycleGAN in both metrics, performs comparably to CUT and FastCUT in terms of JS-MF length, but excels in terms of JS-MF diameter.

Additionally, I compute the total number of myelin fiber segments within each sample image (MF-segments number) and the percentage of area occupied by the

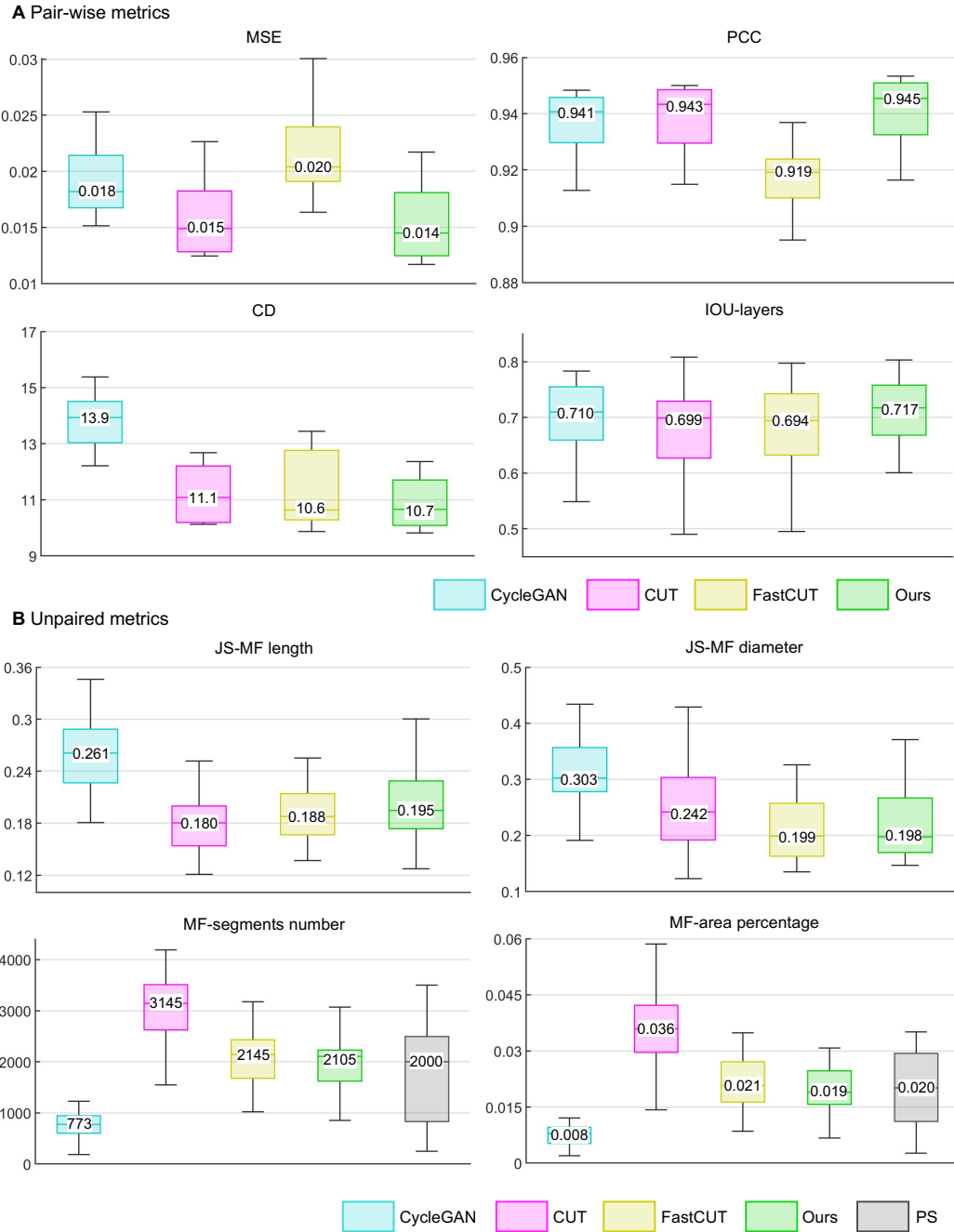


Figure 4-8: Quantitative comparisons with baseline methods. (A) Box plots illustrate four pair-wise metrics, each represented by a unique color denoting the compared methods. The embedded numbers within the plots represent the median values for each metric. (B) Box plots present four unpaired metrics in five colors, with an additional gray color representing the metrics derived from PS images.

fiber segments (MF-area percentage) based on the extracted myelin fiber segments. These two metrics are used to analyze the distributions across the entire sample group, regardless of staining quality, to capture the variability of each method. my method exhibits metric distributions similar to FastCUT, showcasing comparable performance to PS images. Notably, my method demonstrates the closest median value to that of the PS images while reducing variations. Overall, both the paired and unpaired quantitative metrics presented in this section demonstrate the superior performance of my method compared to other baseline methods.

4.4.2 Digital staining enhances mesoscopic brain structures and provides high staining uniformity

I present the ability of my DS technique to preserve the mesoscopic brain structures and achieve uniform staining of cerebral cortex sections from post-mortem human brains. I use two groups of PS imaging results as comparative references: one group consists of WSIs of well-stained sections, and the other group consists of WSIs of less-ideally-stained sections.

In Fig. 4.9A, I present the OCT-SC, DS, and well-stained PS images of adjacent sections from the human cerebral cortex, arranged from left to right. The DS images show that my technique can accurately capture various brain structures that match the PS images, such as cortical layers, myelin fibers, and vessel blobs. The DS and PS images have similar contrast, with white matter (WM) regions appearing as dark brown or black and gray matter (GM) regions appearing as white, while the OCT-SC image has the opposite contrast. my DS model can reliably learn this general inverse mapping between OCT-SC and PS images.

In the zoom-in regions, I present the images on different types of cortex regions, including gyral crest regions marked as 1 and 3 and sulcus regions marked as 2 and 4, from the three modalities: OCT-SC, DS and PS. In region 1, the structures of myelin

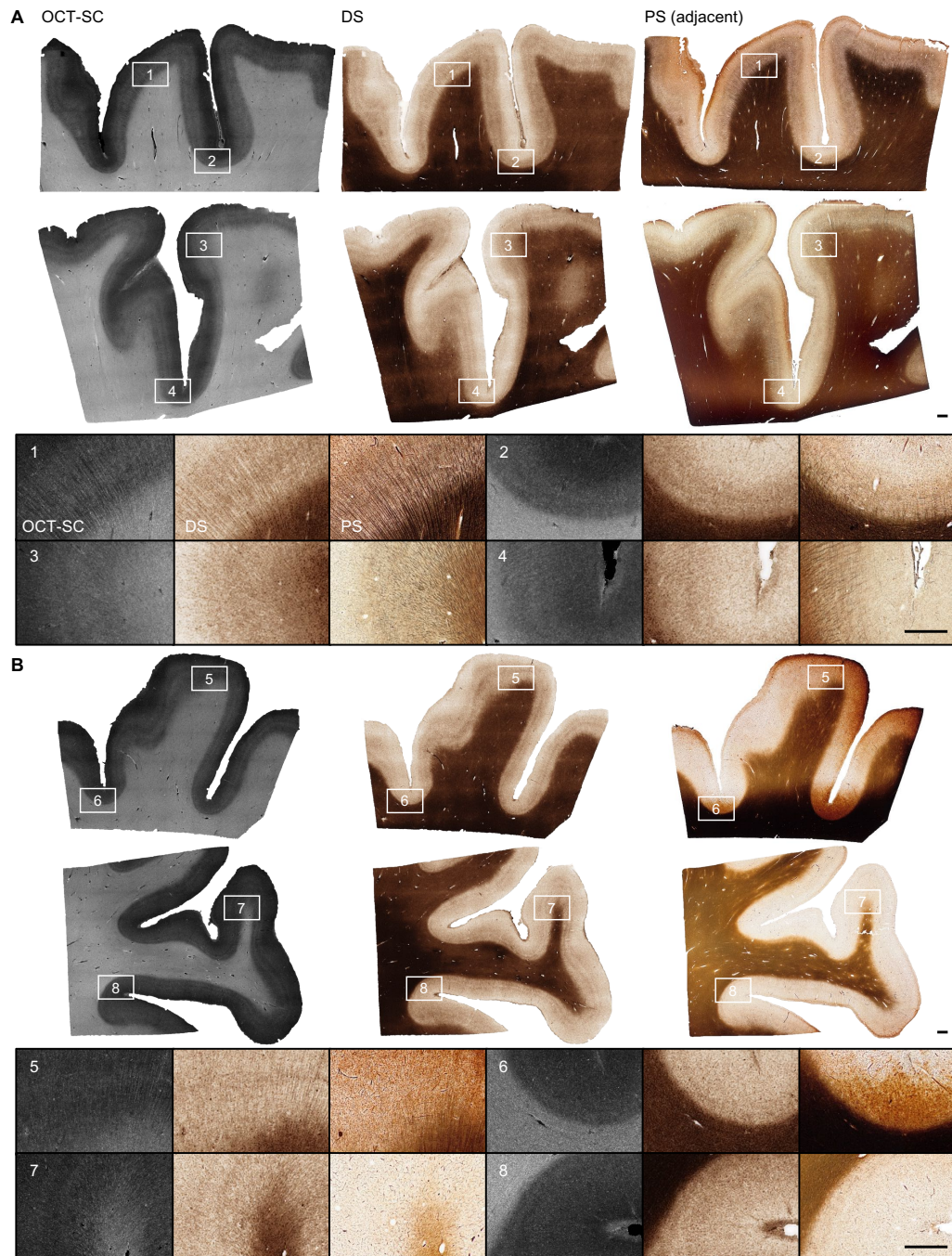


Figure 4-9: DS results on OCT-SC of tissue slices and comparisons with PS images. Cases include (A): ideal staining samples; (B): non-uniform staining and understanding samples. ROI 1, 3, 5, 7 are gyral crest regions and 2, 4, 6, 8 are sulcus regions. Scale bars are 1 mm.

fiber bundles at scales of about 10-20 μm are shown as dark brown tubular features in both DS and PS images, especially in the GM region. By comparing OCT-SC and DS images, I can see that the image content is consistent, which indicates that the ability of resolving fine features is primarily limited by the input OCT-SC data. Despite the limitations of resolution and speckle noise in the OCT data, the orientation of fiber bundle traces and the intensity distribution according to cortical layers can still be discerned in the DS results. Similar patterns are also evident in zoom-in regions 3 and 4, where the local intensity variation is visible in the GM regions, although the fiber bundles are less distinct in OCT-SC and DS images than in the PS images. In region 2, the cortical layer III, layer IV+V, and layer VI are clearly distinguished by the white, light brown, and dark brown bands, respectively. The black line structure near the top of the PS image indicates smaller vessels with glial cells, which are also visible in the DS image at the same locations. The zoom-in regions 1, 2 and 3 show small white blob or tubeness features that represent the vessel structures, especially in the WM regions. These visualizations demonstrate that my DS model can faithfully reveal $\sim 20 \mu\text{m}$ scale brain structures.

A major advantage of DS over PS is stain uniformity. To demonstrate this, I present three types of images in Fig. 4-9B from the less-ideal PS group that comprises most of my PS data. One inherent limitation of traditional histological staining is the variability across different sample regions and experiments. Despite my careful sample preparation and staining procedures, the staining result is influenced by many confounding factors of the chemical reaction and uniformity of the staining quality is difficult to ensure. In Fig. 4-9B, the rightmost column of the first row shows a PS example with over- and non-uniform staining (in particular along the vertical directions); the second row shows a PS example with under-staining.

I select two gyral crest regions (marked as 5 and 7) and two sulcus regions (marked

as 6 and 8) to zoom in. The PS images in regions 5 and 6 are over-stained, while the PS images in regions 7 and 8 are under-stained. In region 5, the DS and OCT-SC images show clear ridges corresponding to cortical layer V, but the PS image shows a dark brown shade due to over-staining. In region 6, which is a sulcus region with less visible cortical layers, the DS image shows a clear boundary between WM and GM regions, but the PS image shows an ambiguous boundary. Small vessel blobs are also more visible in the DS image than in the PS image. In region 7, which is a gyral crest region, the DS image shows dark ridge features corresponding to cortical layer IV and V, but the PS image does not show these features due to under-staining.

Fig. 4-10 provides additional examples of my results obtained from cortex tissue section samples. The figure includes four WSIs including OCT-SC, DS, and PS images (adjacent sections) arranged from left to right. Within each sample, I have selected two smaller region-of-interest (ROI) areas to highlight finer details. Notably, the color contrasts remain consistent, and the preservation of vessel structures is evident in my DS result, indicated by small white blobs. It is worth noting that the cortical layer separation appears to be more pronounced in the DS images compared to the PS modality, primarily due to reduced staining variability.

The first row of the figure displays one of the relatively ideally stained samples. However, in the remaining samples, I observe instances of under-staining, over-staining, or non-uniform staining patterns in the PS images. On the whole, I observe comparable visualization of fine features and superior staining consistency across samples and ROIs in the DS images compared to the PS images. These additional examples align with the results presented in the main manuscript, further supporting my findings.

The superior stain uniformity demonstrated by my DS results across different sections is enabled by the SC calculation that extracts a normalized quantity that

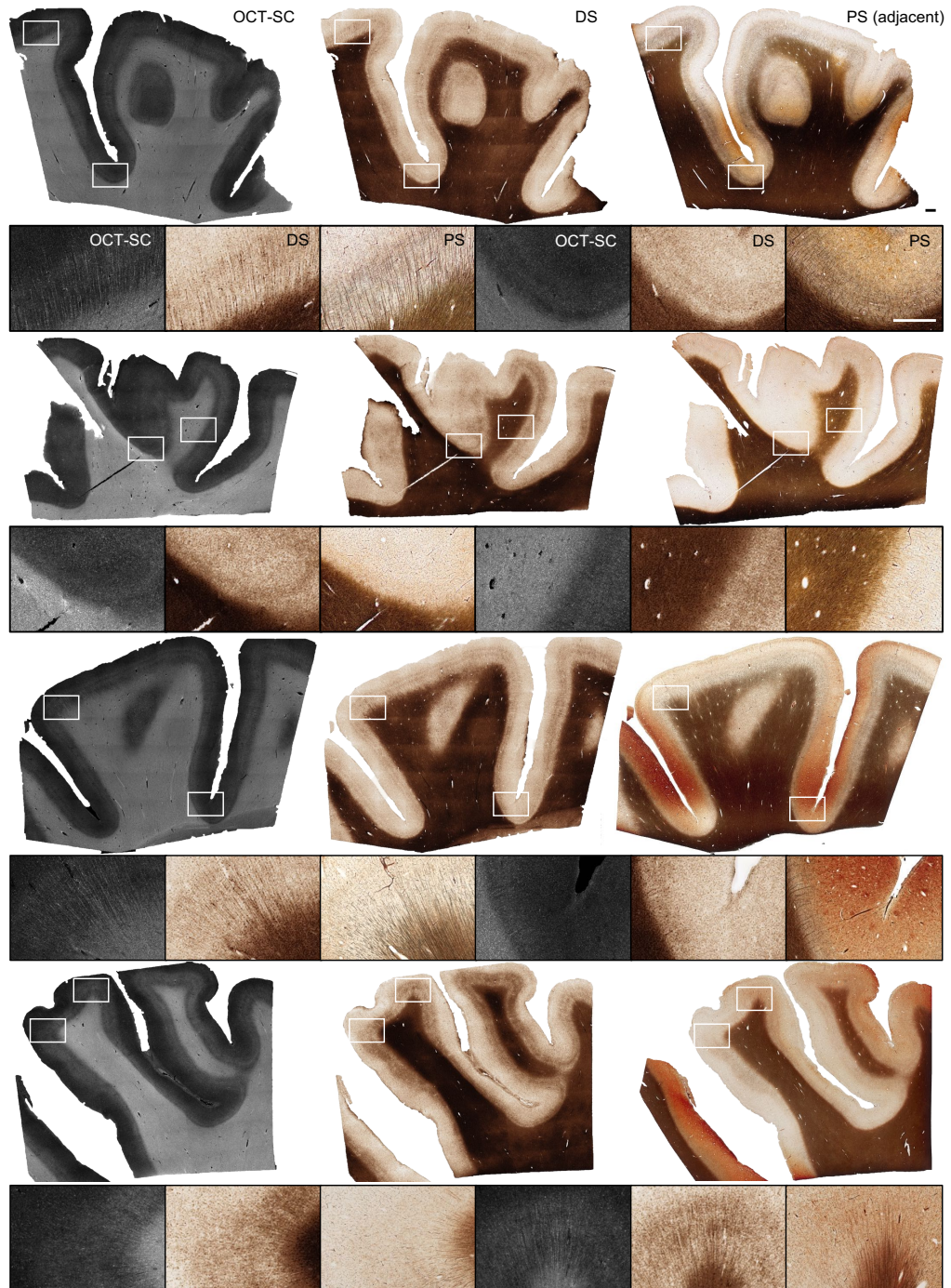


Figure 4-10: Additional visualization of DS results. Cases include: Row 1: ideal staining; Row 2: under-staining sample; Row 3: over-staining sample; Row 4: non-uniform staining. Scale bars are 1 mm.

reflects the intrinsic property of the brain tissue. However, my current SC curve fitting model reduces the spatial resolution (lateral: 6 μm raw OCT measurement, 12 μm fitted SC map; axial: 6 μm raw OCT measurement, 150 μm fitted SC map), which limits the ability to resolve fine fiber structures. I aim to overcome this limitation in my future work.

4.5 Summary

In summary, I present a novel deep learning technique for DS of OCT images for large-scale human brain imaging. I exploit the side information provided by the content similarity between the adjacent sections as well as a biophysical module to guide learning with weakly paired image dataset. my DS framework augments unpaired image translation with pseudo supervised learning and unsupervised cross-modality image registration. Overall, my method achieves superior performance than fully unsupervised baseline methods. The semi-supervised DS framework exhibits extensive applicability to other bioimaging modalities and applications, particularly in cases where paired images are limited.

My method allows direct visualization of important mesoscopic brain features, such as laminar structure of the cerebral cortex with contrast that closely resembles Gallyas-silver staining. my method has shown an important advantages over traditional PS - reducing staining variability. my method also improves the interpretability of the label-free OCT modality for brain imaging.

Chapter 5

Pilot study of DS-OCT for the application of enhancing 3D histological brain imaging and neuropathology

5.1 Background

The human brain consists of billions of neurons that form intricate connections and networks that underlie the complex functions. To gain new insights into the brain, major efforts have recently been made to develop multiscale imaging technologies for visualizing anatomical structures with microscopic resolution across cubic centimeters of tissue. The most widely used techniques for visualizing neuronal structures are based on histological staining. Gallyas silver staining is used to characterize myelin content and neuronal structures, as well as to identify pathological features of neurodegenerative diseases in human brain tissue [92, 95]. To create a high-resolution 3D model of the cytoarchitecture, the BigBrain project reconstructed a whole human brain with more than 7000 histological sections, which involves slicing the tissue into 20 μm thick sections, staining with silver to reveal cellular and fiber structures, and registering the slices in 3D. However, these histological staining processes are generally complex, labor-intensive, time-consuming, and prone to experimental error and staining variability. Furthermore, the slicing, mounting, dehydration, and staining inevitably cause tissue damage and slice-specific distortion, which severely limits the accuracy of 3D alignment and reconstruction of structures at the micron scale [96, 97].

Therefore, there is a growing need for developing 3D pathology imaging techniques, especially label-free techniques that can provide high-resolution 3D visualizations of brain tissues with minimal tissue damage and distortion, and reduce the need for physical staining (PS) [98, 46, 99, 100].

5.2 Overview

In the previous chapter, my work established a novel semi-supervised digital staining method that learns from weakly-paired image data and achieves improvements over previous baseline methods. My work also demonstrated the advantages of staining uniformity of my DS-OCT technique compared to traditional PS modality.

In this chapter, as a pilot study of the potential applications of DS-OCT to practical pathological interpretation, I show that DS enhances contrast across cortical layer boundaries and can consistently differentiate cortical layers IV/V/VI. In addition, I show a 3D-rendered volumetric DS result on a cubic centimeter-scale tissue block that was not used for training the DS model. The result shows geometry-preserving 3D staining on large-scale brain tissue and visualization of vessel structure in the white matter region. Finally, I showcase a pilot study on the generalization performance of my method - I apply the DS model trained on cortex regions to samples from other anatomical regions acquired from different OCT setups.

5.3 Image analysis

The layer differentiation analysis is primarily performed using the open source ImageJ software package. The line profiles are computed by selecting the rectangular region in the center region of interest (ROI) and aggregating the intensity value along the horizontal direction. Those profiles are then normalized to $[0, 1]$ by their individual value range for visual comparisons. The cortical layer boundaries are manually

annotated by identifying the local maxima and edges according to [101, 102]. The layer segmentation on the larger ROI is performed by manual annotation on layer IV/V/VI. I used the built-in local thickness estimation function to generate the local thickness map and calculated the box plot for the thickness distribution using Matlab. Two Gyral crest ROIs and one Sulcus ROI are manually selected.

5.4 Results

5.4.1 Digital staining enables reliable cortical layer differentiation and layer thickness quantification

I demonstrate the capability of OCT DS to reliably distinguish cortical layers with comparable or even better sensitivity than PS, thanks to the uniform DS quality as discussed before. I identify cortical layers IV, V and VI by the displayed fiber density [101, 102], since these layers are more prominent than layers I, II and III in most of my samples. I also show how the layer thickness can be consistently quantified in my DS images.

Fig.5.1A shows the WSIs of the DS result and the reference PS of an adjacent brain slice. The DS image clearly reveals the curved double-band structures around the WM region, which are stained in dark brown. These features indicate higher myelin fiber density that are characteristic in cortical layer IV and V [102]. Consistent image contrast variations for the laminar structures are observed in the DS result. In contrast, the double-band structures are less visible around some of the gyral regions and the contrast is less distinct in the PS image. Fig.5.1B shows zoom-ins from a gyral crest region and a sulcus region of the three modalities, corresponding to the regions marked by the green box and red box in Fig.5.1A respectively. The OCT-SC and DS images have a strong correlation in their intensity variations. The DS image consistently shows the double-band features in the GM region, while the PS image

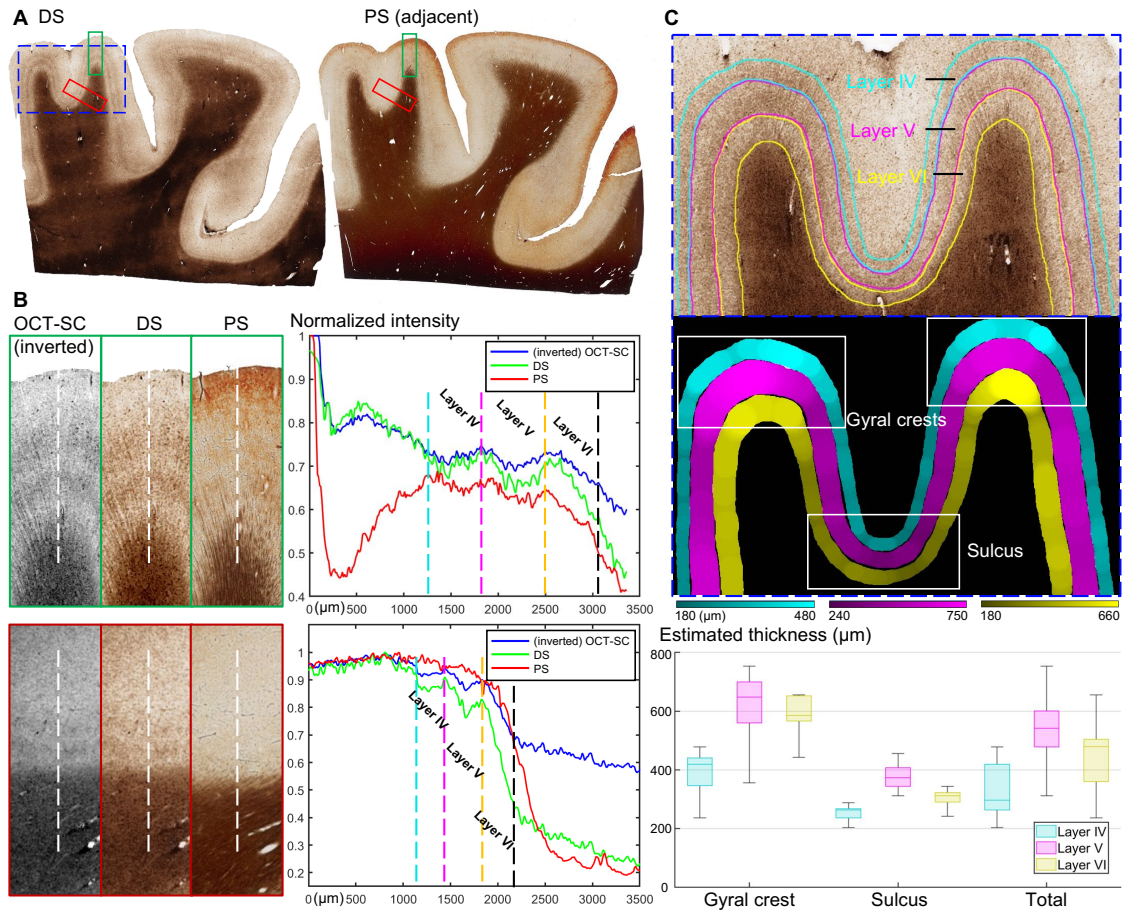


Figure 5-1: Comparisons results of layer differentiation and thickness estimation in DS results. (A) The DS and PS WSIs from a cortex tissue section. (B) Zoom-in ROIs of inverted OCT-SC, DS and PS modalities marked in green and red boxes in (A) and normalized intensity profiles aggregates along white dotted lines. (C) Manually annotated layers IV/V/VI labeled in three colors and estimated local thickness. Statistics of thickness are visualized in box plot and grouped by gyral crest and sulcus regions. ROI is the zoom-in of dotted blue box from (A).

often fails to reveal them due to over- or under-staining.

Next, I demonstrate the improved contrast between cortical layers in DS by plotting the average intensity (across the three color channels) along the white dotted lines in Fig.5-1B. The right panel shows the normalized profiles over a 3.5-mm depth range, where blue, green and red represent OCT-SC, DS and PS modalities, respectively. I manually marked the boundaries of layer IV, V and VI with dotted vertical

lines in four different colors. In both gyrus and sulcus regions, the DS profiles show the highest contrast (measured by the difference between the maximum and minimum values) in layer IV and V among the three modalities, which facilitates identifying the layer boundaries. When comparing OCT-SC and PS with DS, the DS enhances the intensity variations at the boundary between layer IV and V. This reduces any confusion when distinguishing between these two layers. Comparing the profiles between OCT-SC and DS in different layers suggests that my DS model works beyond my approximate linear biophysical model and increases the local contrast by a nonlinear mapping function expressed by my neural network.

In Fig.5-1C, I further demonstrate straightforward segmentation and thickness quantification of cortical layers IV, V and VI using my DS result, which can provide valuable information for many neuropathological studies. The top panel shows the zoom-in region of the dotted blue box in Fig.5-1A, where I manually labeled the boundaries of the three cortical layers. I estimated the layer thicknesses from the binary mask obtained from cortical layer segmentation using an algorithm from my previous work. I chose two gyral crest regions and a sulcus region indicated by the white boxes in the binary mask image. The bottom panel displays the box plot of the local layer thickness statistics in gyrus and sulcus regions. I observed a similar pattern of variation in layer thickness for layer IV, V and VI in the sulcus, gyrus and the entire cortical regions. The median local thickness of layer IV, V and VI were 300 μm , 540 μm and 480 μm respectively. I also observed a significant reduction in layer thickness in all three layers in the sulcus regions compared to the gyrus regions, in agreement with the literature. The median thickness of layer IV, V and VI were 410 μm , 630 μm and 580 μm respectively in the gyrus regions, and were 250 μm , 370 μm and 310 μm respectively in the sulcus regions.

Fig. 5-2 provides additional examples highlighting the improved differentiation of

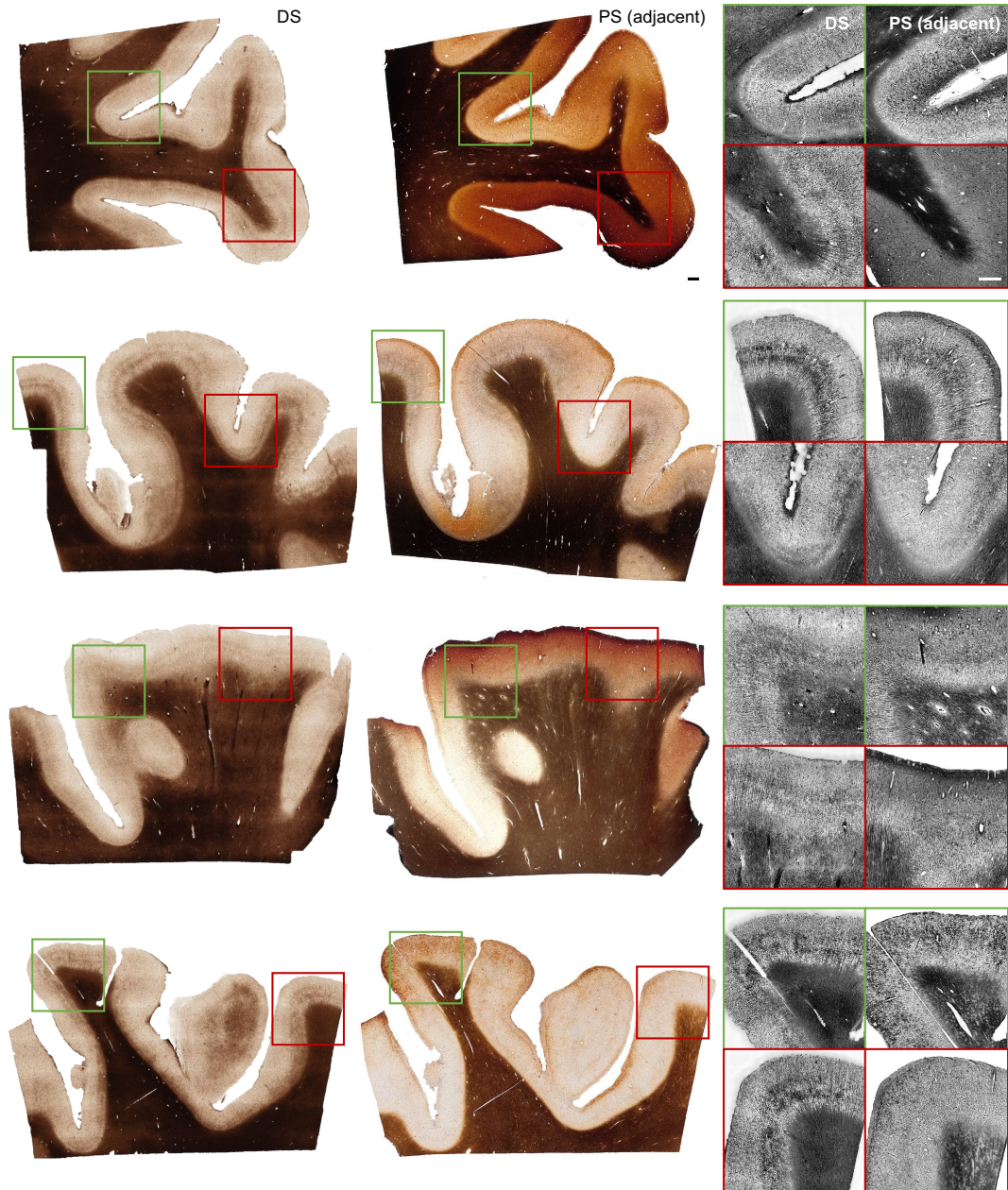


Figure 5.2: Additional examples of improved layer differentiation in DS results. The DS and PS WSIs from four additional cortex tissue samples are shown on the left panel. Contrast enhanced grayscale images are shown in zoom-in ROIs. For these ROIs, The DS and PS ROIs are shown in order from left to right, while different ROIs are shown from top to bottom and marked by green and red boxes. Scale bars are 1 mm.

cortical layers in my DS results from cortex tissue section samples. The figure includes four examples of WSIs of DS and PS images (adjacent sections). Within each sample, two smaller ROIs are selected. To enhance the contrast of cortical layers, the same image enhancement process, utilizing contrast-limited adaptive histogram equalization (CLAHE) and conversion to grayscale, is applied to each ROI. Across all the examples, I consistently observe distinct double band features present in the zoomed-in ROIs of the DS results, whereas only a portion of these features is visible in the PS results. Moreover, within the same sample, the DS images exhibit uniform double band features in both sulcus and gyrus regions, whereas the PS images only exhibit some of these features. It is important to note that the PS modality demonstrates instances of both over-staining and under-staining, leading to inferior visualization of cortical layers. Consequently, the DS images facilitate easier differentiation of cortical layers compared to the PS modality due to reduced staining variability. These additional examples serve as further evidence supporting my argument that DS reduces staining variability, enabling improved differentiation of cortical layers.

5.4.2 Volumetric digital staining on cubic centimeter-scale brain tissue

Next, I showcase volumetric staining on cubic centimeter-scale brain tissue enabled by my technique that combines S-OCT and DS. My technique significantly reduces tissue distortion and misalignment during the 3D reconstruction process suffered by the traditional 3D pathology technique. I demonstrate 3D DS on a 4 cm \times 5 cm \times 1.2 cm brain tissue block that was not used for training my DS model. I show that my method can preserve the intricate 3D brain structures in both GM and WM regions. Moreover, I visualize the 3D vessel network in the WM.

In Fig. 5-3A, I present a 3D visualization of the DS output on the whole tissue block in the top panel. The DS model takes as input a z-stack of around a hundred slices of OCT-SC images. Each OCT-SC slice, which has a size of 4 cm \times 5 cm, is

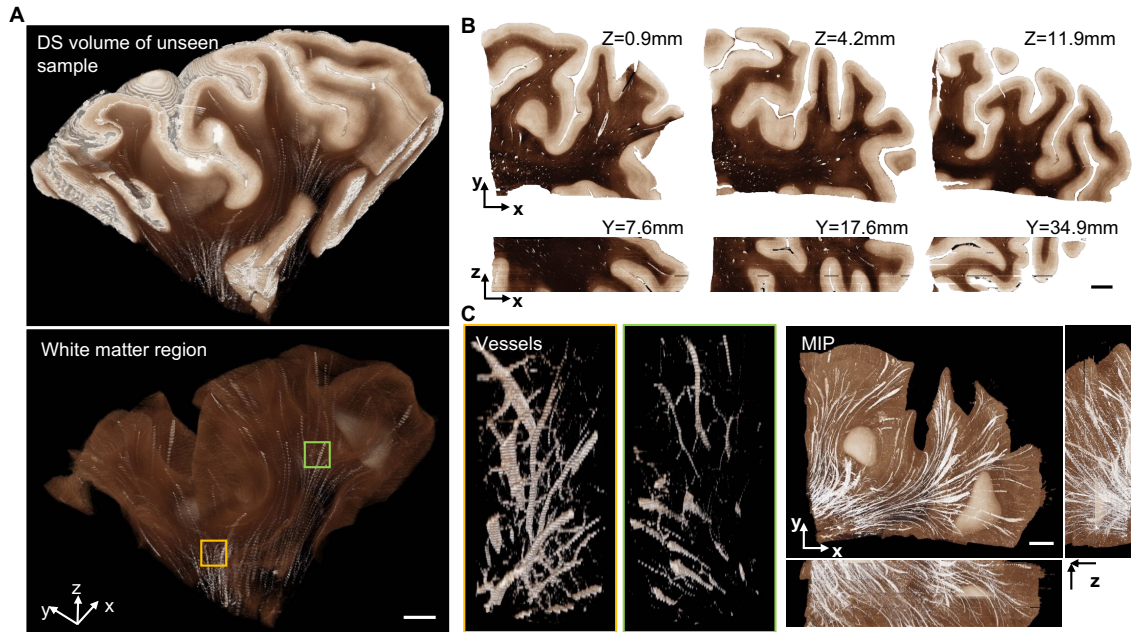


Figure 5-3: 3D visualization and cross-sections views of the DS results on a large unseen tissue block. (A) The DS output images are stacked along the z-axis to render the whole digitally stained volume as well as segmented WM regions. (B) Orthogonal cross-sectional views of the DS volume. (C) Two zoom-in regions of vessel structures in yellow and green boxes from (A) are shown on the left. Three orthogonal maximum intensity projections (MIP) of the DS volume are shown on the right. Scale bars are 5 mm.

processed separately and fed to the DS model. The DS output images are then directly stacked along the z-axis to create the digitally stained volume. Consistent with the 2D results, the 3D DS volume generates white and dark-brown colors that correspond to GM and WM regions respectively. I can also observe a smooth transition of these GM and WM boundaries along the z direction, which reflects the preservation of 3D geometries of the brain structures. In Fig. 5-3B, I display several orthogonal cross-sectional views of the DS volume. The overall color tone and contrast variations match with the 2D results in Fig. 2-12. Small white blobs and tubes within the WM region indicate the blood vessels. These results are consistent with 2D DS results that have been verified with PS references, and partly confirm the generalization

ability of my DS model on unseen large-scale brain samples. Moreover, the X-Z cross section also shows several continuous features along the depth, such as intricate brain folding structures, double-band cortical layers, and small tubular vessels. This again illustrates the 3D geometry preservation feature of my DS technique.

To further illustrate the ability of my DS technique to preserve the 3D geometry of mesoscale brain structures, I present a 3D visualization of a centimeter-scale vessel network, which is not visible in 2D PS images. Besides the GM and WM contrast, my DS volume also shows several continuous white tubular structures corresponding to blood vessels in the top panel of Fig. 5-3A. In the bottom panel of Fig. 5-3A, I show the segmented DS volume displaying only the WM region, where the white tubular structures are more prominent and not masked by the GM. In Fig. 5-3C, I highlight two regions in yellow and green boxes. The vessels in those regions are rendered with more transparency and reveal the branching and connectivity of the vessel network. On the right panel of Fig. 5-3C, three orthogonal maximum intensity projections (MIP) of the DS volume further demonstrate the preservation of the 3D vessel structures. I note that the axial continuity of my DS volume is currently limited by the axial resolution (150 μm) imposed by my SC fitting model, which I aim to improve in the future.

5.4.3 Generalization to unseen anatomical regions

To further demonstrate the generalization capability of my trained DS model, I conducted a pilot study on different anatomical regions and S-OCT setups that were not seen during training. I used the same fitting model to generate the OCT-SC image in Fig. 5-4, which shows a sample from the hippocampus region acquired by a different S-OCT setup. Since my SC fitting model extracts an intrinsic tissue property and is relatively insensitive to variations in hardware platforms and sample conditions, it ensures the robustness of my DS method. The DS image is inferred by directly

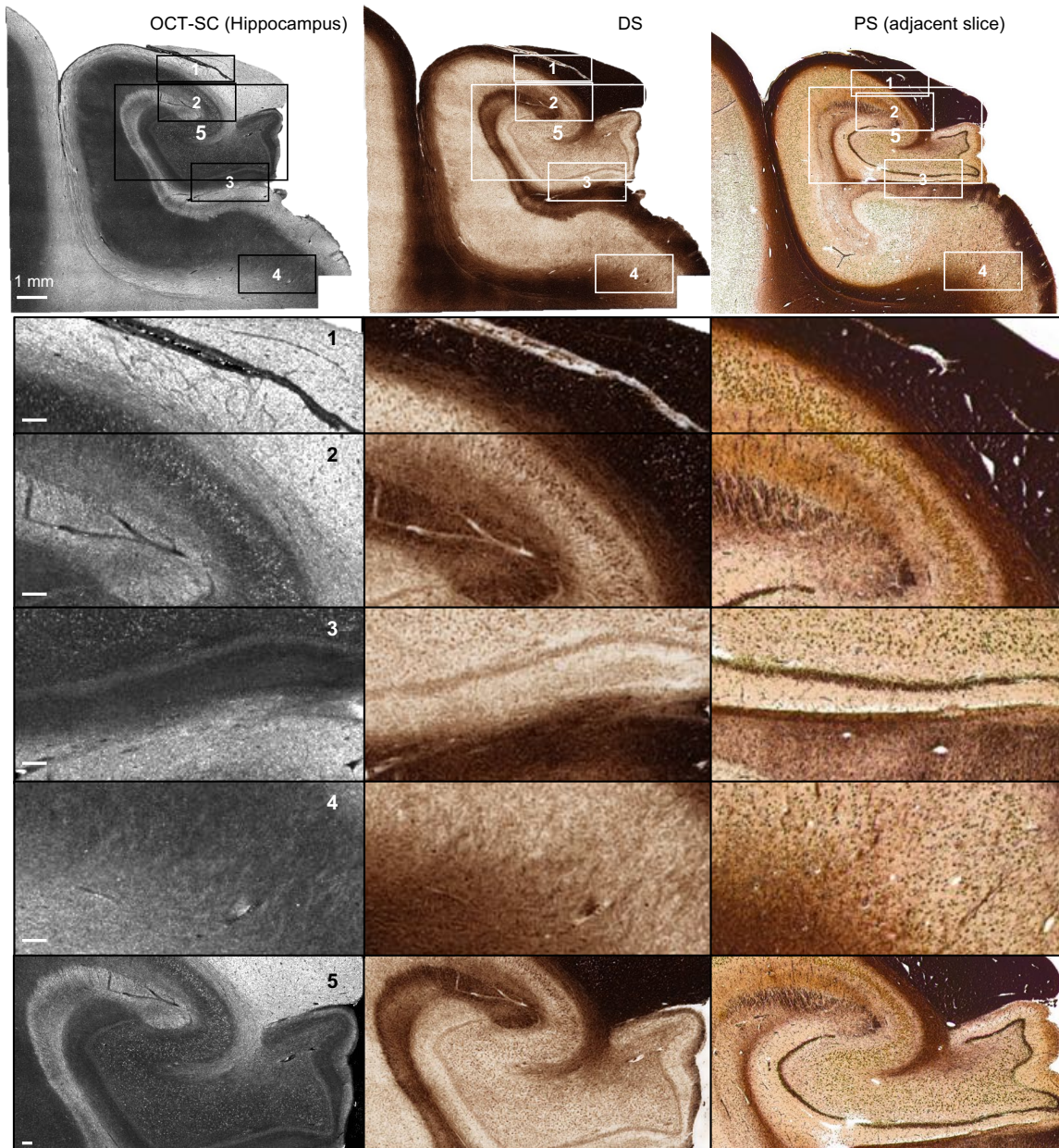


Figure 5.4: DS-OCT generalization performance on a hippocampus tissue slice. Examples of OCT-SC, DS and PS images (of adjacent sections) on one sample from Hippocampus region are shown. Each zoom-in ROIs correspond to numbered boxes in WSIs. Scale bars are 1 mm.

inputting the OCT-SC to the previously trained model. Fig. 5.4 shows the OCT-SC and DS images, and the reference PS image of an adjacent section from left to right.

I roughly aligned the field of views of the DS and PS images using a rigid transformation. The comparison between DS and PS shows that my model can successfully transform the image contrast while preserving the fine features in the OCT measurement. In zoom-in region 1 and 2, I can see the same layer structures in black, dark brown and white in the DS and PS images. In region 3, despite the distortion of the field of view, I can still identify the dark ridge feature and the layers in both modalities. The dark ridge line also appears in the bottom right corner of region 2 in both DS and PS. Moreover, I notice small features that look like dark spots in region 1, 2 and 3 in both modalities, which are likely indicative to the local distribution of neuron cell density.

As in the previous case of the cortical regions, the results of this initial generalization test are also affected by the limitations of the S-OCT systems, such as image stitching artifacts, uneven tissue slicing, speckle noise, and limited lateral and axial resolution due to the SC model fitting. These issues can be seen in more examples from other anatomical regions in Fig. 5-5. I plan to conduct a more comprehensive study in the future to fully investigate the generalization capability of my current DS framework.

In Fig. 5-5, I provide additional examples that illustrate the generalization capabilities of my DS model to other anatomical regions and OCT imaging setups. The figure showcases four distinct examples, each displaying OCT-SC, DS, and PS images of adjacent sections, acquired under different experimental conditions and S-OCT setups that differ from those presented in Chapter 4.

In Fig. 5-5A, I present the results obtained from a cerebellum sample using three modalities. Due to the presence of significant speckle noise in the OCT-SC image, a preprocessing step involving outlier removal and strong averaging was applied, resulting in an over-smooth appearance. Consequently, my DS result also exhibits

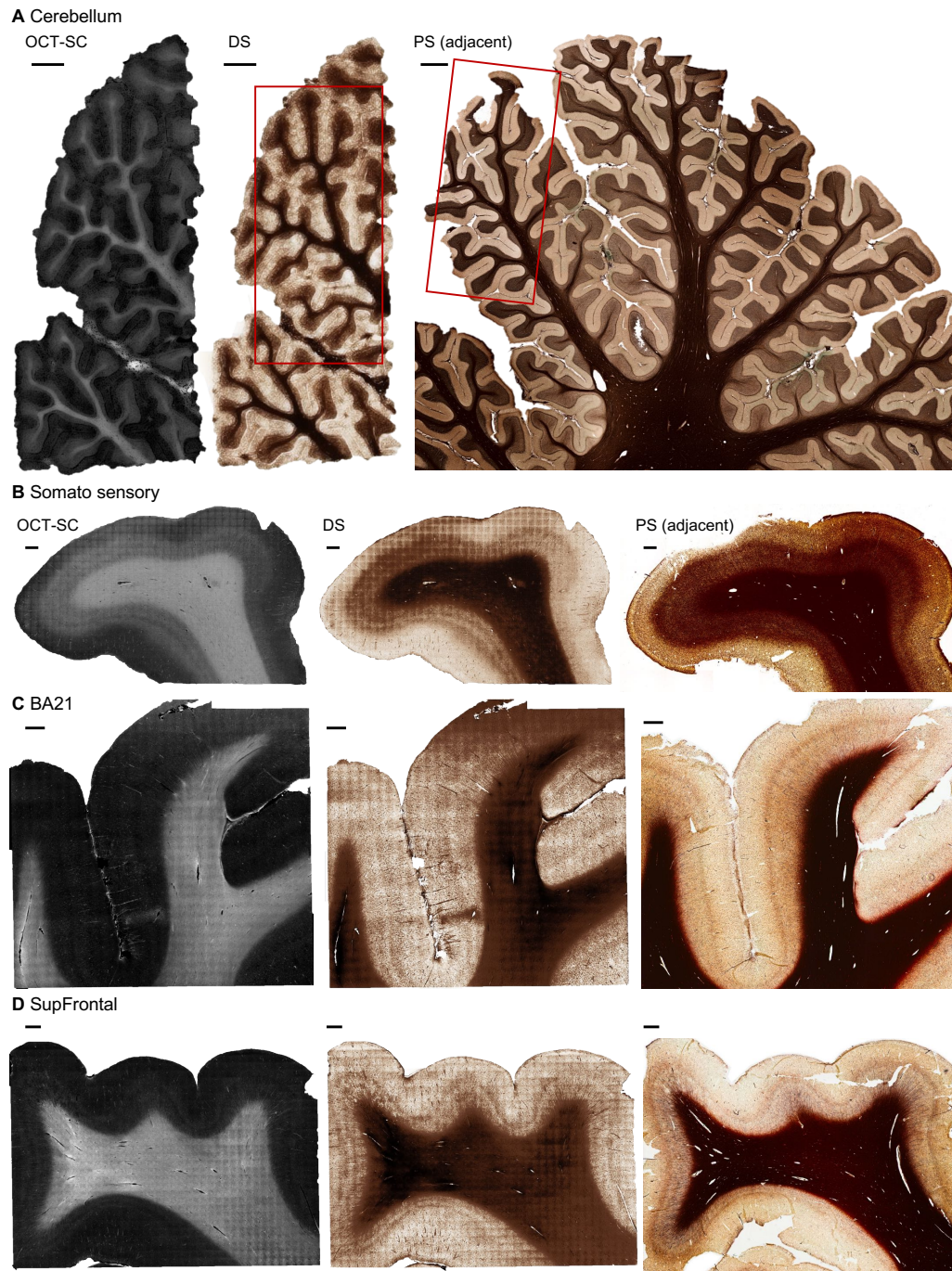


Figure 5-5: Additional examples of DS-OCT generalization performance. (A-D) tissue sections from four anatomical regions: cerebellum, somatosensory, BA21, and SupFrontal, respectively. In (A), red boxes are coarsely registered ROIs. Scale bars are 1 mm.

smoothing artifacts. Despite the reduced resolution, I observe similarities in color and layer contrast between DS and PS images, which differ noticeably from samples obtained from the cortex region. Within the sample, I identify three distinct layers stained with light-brown, dark-brown, and black, progressing from the outer to inner regions. Manual annotation of corresponding regions of interest (ROIs) is represented by two red boxes, highlighting the distortion in sample geometry. For the remaining samples with less distortion, coarsely registered ROIs are displayed side by side. Fig. 5.5B, C, and D depict three samples from different anatomical brain regions: Somatosensory, BA21, and SupFrontal, respectively. The OCT-SC modality demonstrates some degree of contrast variability due to system imperfections and strong stitching artifacts. Consequently, the DS results also exhibit stitching artifacts and less staining uniformity. Given the presence of these artifacts, visualizing myelin fibers or fine features in the DS results becomes challenging. However, coarse cortical layers can still be observed in the DS images as evidenced by dark ridges. Within the gray matter region, DS successfully differentiates two cortical layers in Fig. 5.5B and D, and a single cortical layer in Fig. 5.5C, aligning well with the PS results. In summary, it is evident that the performance of my current DS-OCT model is influenced by speckle noise and stitching artifacts. These results highlight the dependence of my DS-OCT model on the imaging quality of the S-OCT setup and data processing pipeline.

5.5 Summary

In this chapter, I show further advantages of DS-OCT over traditional PS, such as better cortical layer separation, preserving complex brain 3D geometry and facilitating volume generation across cubic centimeters of tissue. I present results of pilot study on the application of my DS-OCT technique for neuropathology image analysis.

I have shown the reduced staining variability of DS is important in uniform visualization of cortical layers and the DS results are useful for layer thickness estimation and quantification. On entirely unseen volume samples, my method shows good generalization performance and direct visualization of main 3D blood vessel network in the white matter, which cannot be achieved in traditional histological staining. To further test the practicality of the trained DS model, I generalize the model to other OCT setups and samples from other anatomical regions. My results have significant implications for quantitative volumetric neuropathology.

However, my method also faces some limitations that originated from my current S-OCT system, such as artifacts from image stitching, uneven tissue sectioning, speckle noise, and limited lateral and axial resolution due to the SC model fitting. Although my technique is sensitive to fiber structures in the gray matter, the speckle noise and limited resolution resulted in discontinuities and grainy artifacts in the DS results.

Chapter 6

Conclusions

To conclude, deep learning has shown great potential in bridging the gap between label-free imaging and fluorescence and histology imaging modalities. In my thesis, I have developed several innovative solutions to address the challenges associated with incorporating deep learning-generated digital stains into label-free imaging techniques. I have explored both supervised and semi-supervised approaches to develop innovative DS frameworks. The outcomes of my work have showcased their application in two significant areas: immunofluorescence single-cell imaging and 3D histological brain imaging.

Through this exploration, my goal is to shed light on the transformative impact of digital staining in the field of biomedical computational imaging. By bridging the gap between traditional imaging techniques and deep learning, I anticipate that digital staining will revolutionize various scientific disciplines by enabling new avenues of biomedical research and analysis.

6.1 Multiplexed digital fluorescence labeling for single-cell cytometry

By integrating DS techniques with multi-contrast microscopy, I have uncovered the potential to greatly enhance the efficiency of single-cell imaging cytometry and phenotyping. This advancement opens up avenues for high-throughput analysis and characterization of individual cells, potentially leading to a deeper understanding of

cellular behavior and biological processes.

6.1.1 Summary

In summary, I have reported a label-free imaging cytometry technique that multiplexes six IF labels in-parallel with high accuracy via DL models. I have validated the fluorescence predictions by comparing them to the ground-truth IF images. In addition, I have conducted imaging cytometry studies on several quantitative morphological metrics on subcellular structures and phenotyping of cell proliferation. Finally, the specificity of the DL model is assessed by visualizing the saliency map at the single cell level across different staining and fixation conditions. With this unique combination of new capabilities, this new framework may find wide applications in image-based cytometry, in particular for high-content screening and analysis.

6.1.2 Discussion

Label-free, DL-augmented method of cell-morphology profiling is data-driven and ultimately relies on the rich information content in the images. Our LED-array reflectance microscopy enables multi-contrast imaging (i.e. angle-dependent dark-field and drDPC) by detecting the angled-dependent backscattering signals by a programmable LED array without any mechanical moving parts. The superior sensitivity in detecting subtle structures using backscattering than transmission-microscopy is well documented [51]. The superb sensitivity of backscattering-based method has been demonstrated in a variety of techniques, such as partial wave spectroscopy [103], confocal light absorption and scattering spectroscopic microscopy [104], confocal reflectance quantitative phase microscopy [105], and spatial-domain low-coherence quantitative phase microscopy [106]. In addition, the angle-dependent measurement has been used to measure characteristic structural length scale [55] and to enable 3D reconstruction of the refractive index distribution [107]. Leveraging angle-dependent

reflectance signal, I outperformed the state-of-the-art for predicting multiple subcellular components. By switching to a higher numerical aperture (NA) objective lens, the prediction performance of my DL model to subcellular structures, such as those missed in my actin label predictions, may be further improved. Recently, the LED array microscopy has also been extensively explored in transmission that allows sampling the low spatial frequency components in the Fourier space [107]. In addition, backscattering spectroscopic techniques further enable characterization of ultrastructural phenotypes with sensitivity down to nm-length scale [108]. A potential future improvement of my imaging system is to incorporate additional transmission and multispectral LED-array illumination to fully exploit the angle- and wavelength-dependent scattering contrast with a single objective lens by versatile illumination engineering.

Cell morphological features are effective phenotypes for different disease states and environmental influences. This phenomenon is well described and practiced in pathology and cell biology. Nuclear condensation, enlargement, and increased NRC are ubiquitous hallmarks of cancers [54, 55]. Cell morphology is distinct for different cell types, which is often denoted in their terminology (i.e. astrocytes, macrophage, squamous, and columnar cells, etc.), and stem cells change structures along separate differentiation paths [60]. It has been shown that cell morphological changes can be directly associated with changes of morphogenic gene expressions [109], and comprehensive morphological profiling can be used to detect genetic functions [110]. Meanwhile, it has been shown that DL techniques can holistically capture complex structural features for classification. This has found broad applications in detecting cell types [60, 61, 111], cell states [62, 63, 64, 65, 69], drug response [66], and stem cell lineage [67]. By fully leveraging the label-free and high multiplexing nature of my technique, it can potentially generate significant impacts in imaging cytom-

etry by offering unprecedented information content and discovering new compound morphological features necessitating multiplexed fluorescence readouts.

6.1.3 Future work

One limitation of my current work is that it is based on fixed cells that does not allow longitudinal imaging. This can be overcome by using fluorescent reporter cell lines or live cell dyes to provide the fluorescence ground-truth [20] and enable dynamic observation. The additional temporal dimension may further improve the model's sensitivity in cell phenotyping and discover new label-free features by incorporating the information about the cell dynamics [67, 69, 25]. Another limitation of the DL framework I used here is that it cannot be generalized to different types of cells. Techniques based on transfer learning [25, 26] and domain adaptation [112] could be investigated in the future work to overcome this limitation. The variations in the prediction accuracy of my DL models are currently evaluated post hoc, and based on pixel-wise and cell-level metrics by comparing the DL predictions and the ground truth. For many biomedical applications, it is beneficial to understand how much error the model may make without knowing the ground truth, i.e. the confidence of the model predictions. Emerging Bayesian DL based uncertainty quantification techniques have proved useful to provide a proxy estimate of the prediction accuracy and quantify the model confidence [10, 113], which can be adapted in the future work.

6.2 Semi-supervised DS-OCT for 3D histological brain imaging

The integration of DS techniques with S-OCT has demonstrated immense promise for high-throughput human brain imaging. This integration may enable comprehensive investigations into the structure and function of the brain, providing valuable insights into neurobiology and contributing to advancements in neuroscience research.

6.2.1 Summary

In summary, I developed a novel semi-supervised learning technique for DS of OCT images for large-scale volumetric visualization and analysis on human brain tissue samples. The technique works by integrating label-free S-OCT imaging and an advanced deep learning DS model. The S-OCT enables imaging of cubic centimeter-scale brain tissues and preserves complex 3D tissue geometry across sections. My semi-supervised learning method bypasses the need for paired unstained and stained images and can achieve high-quality DS using a limited amount of weakly paired image data for model training. My deep learning model is built on an unsupervised CUT model backbone, which is augmented with two auxiliary tasks. The biophysically inspired module reduces the data requirement for model training by exploiting the correlation between the OCT-SC and the OD of Gallyas silver stain. The unsupervised cross-modality image registration module exploits the structural information between the adjacent tissue sections. By working with a fitted tissue property, namely the SC, from the raw OCT measurement as the input to the deep learning model, it greatly enhances the uniformity and generalizability of the DS results. This is highlighted by my volumetric DS result on cubic centimeter-scale brain tissue block and on unseen anatomical regions from different OCT systems. I believe the OCT DS technique is a promising solution for large-scale human brain imaging for comprehension characterization of brain structure across scales. Broadly, I anticipate that my deep learning DS framework can be broadly applied to many other pathology applications.

6.2.2 Discussion

Despite the advantages of DS-OCT, I discuss some of the main limitations that affect the quality of the S-OCT images and the DS method based on them. The first limitation is the speckle noise that arises from coherent scattering in OCT imaging.

The speckles create randomly distributed fine-grained dark or white spots in the OCT and the fitted SC images that do not correspond to any cortical structures. This makes it difficult to reliably visualize and digitally stain small vessels or fine fiber structures. The second is the limited resolution of my OCT system, which is not sufficient to resolve fine structures such as single neurons. I also applied 4×4 -pixel local averaging to suppress the noise in my images before fitting the SC map, which further reduced my lateral resolution to 10-20 μm . Despite the clear layer differentiation, the fibers appear disconnected and the speckles are still visible. The third is the stitching artifacts that cannot be fully corrected in my current dataset and affect the WSI image qualities. The last limitation is the low axial resolution due to the z-stack processing and SC fitting algorithm. This algorithm assumes that the SC is axially homogeneous within a defined fitting depth range and uses a pixel-wise intensity attenuation model to fit the data. However, this assumption may not be valid for some samples with complex structures with heterogeneous optical properties. Moreover, the fitting depth range I used (150 μm) is much larger than the actual sectioning thickness (50 μm), which introduces an averaging effect along the z-axis and further reduces the ability to resolve individual neuronal bodies. As a result, the fitted SC and volumetric DS results have a large z-spacing and appear blurred and discretized along the z-axis. I expect that these issues will likely be overcome by future generations of high-resolution S-OCT systems, improved pre-processing algorithms and more advanced deep learning algorithms.

Despite current limitations, I believe that my semi-supervised learning based DS framework is broadly useful to other bioimaging modalities and applications. The integration of DS techniques with S-OCT has great potential for high-throughput, multiscale human brain imaging. The data generated from this technique could help better understand the meso- and micro-structure of brain tissues and their role in

disease development, and ultimately enhance my knowledge of the brain’s structure and function.

6.2.3 Future work

I propose several potential directions for future work to enhance the performance and applicability of my OCT DS technique. First, one can optimize the system design and processing pipeline of OCT-SC to achieve higher imaging quality. This involves improving the processing algorithms for speckle suppression, stitching correction, and SC fitting model. These improvements could increase the robustness and resolution of my method and enable us to capture finer neuronal structures. Second, I can incorporate multi-modality input such as polarization information into my DS model to increase the imaging sensitivity to birefringence structures, such as fibers. Third, I can develop a more advanced algorithm that can account for the variations of SC along the z-axis and use a smaller fitting depth range to improve the axial resolution.

6.3 General Discussions

While digital staining techniques offer several advantages, there are still general limitations and challenges underlying the practicality and developments of these technologies. On the one hand, obtaining accurate ground truth for training deep learning models in digital staining is a significant challenge. It requires expert biologists and pathologists to physically perform staining experiments, manually collect and validate fluorescently labeled or histologically stained images, which can be subjective and time-consuming. Therefore, developing reliable and standard datasets for training and validating deep learning models is crucial. On the other hand, despite the standard experiment control, staining protocols used in different laboratories can vary, leading to differences in staining patterns. Digital staining methods should be robust enough to handle these variations and adapt to different staining protocols

to provide accurate digital stains. In addition, current single-modality-based digital staining techniques sometimes struggle to accurately replicate the intensity levels and highlight the desired fine details observed in physical staining. This limitation can affect the fidelity of digital stains, making them less reliable for quantitative cellular and pathological analysis.

To address the above challenges, there are a few potential areas of research from the algorithm perspective. First of all, there is a growing need of exploring and advancing the semi-supervised learning strategies to fully utilize the rich information of unpaired or weakly paired images. Obtaining pairs of label-free images and stained images are generally hard in many biomedical imaging modalities, while it is convenient to acquire images of slightly distorted samples or adjacent samples. By exploiting the inverse modeling or mapping from stained images to label-free modality, one can simulate pairs of pixel-aligned image data as augmented supervision. In addition, exploring more advanced and robust cross-modality registration algorithms is also of great importance to correct the sample distortions and account for the intricate sample geometries. Recent advancements in neural fields based image registration is one of the possible improvements [114]. By learning an implicit coordinate mapping, the neural field can be used to model the geometry transformation off-the-grid and provide higher registration accuracy. As a result, refining and advancing semi-supervised learning algorithms allows for easier training of models on larger and more diverse datasets (potentially unpaired or weakly paired) that are more naturally acquired from biologists and pathologists. Integrating semi-supervised methods can improve the efficiency of ‘data collection-training-validation’ cycle of developing digital staining models. The improved trainability of semi-supervised models on large diverse datasets increases the capability to capture the complexity and variability of staining patterns and therefore improve the staining accuracy and robustness.

Second, as much as we desire standardized staining protocols across different laboratories, this is not always the case in reality due to the various specific needs to highlight different stained features. To make digital staining methods adaptive to various staining protocols practiced in different experiments, one possible solution is to utilize the generative abilities of deep learning models to learn a continuous manifold that controls the attributes of the staining patterns to accommodate the researchers' specific needs, such as the concentration of color tone and the degree of over-staining. This may be done by integrating manually graded staining attributes as additional labels for conditional generative learning. As a result, during inference time, the digital staining models can be fed with additional prompts to generate several realistic staining attributes according to the specific staining style required. Learning to decouple the content and style latent manifold in generative image translation is also intriguing since it may allow for simultaneous multi-modality digital staining by preserving the content latent code and manipulating the style latent code [115]. Overall, investigating the latent manifold of semi-supervised models can provide further insights on ways to improve the robustness and adaptability of digital staining models.

Third, integrating digital staining techniques with multi-modality imaging can enhance the sensitivity to structural features of multiple scales and improve the learned digital staining accuracy, fidelity and reliability. Multi-modality label-free imaging provides a more comprehensive and detailed view of tissue samples with respect to diverse biological and physical properties, such as auto-fluorescence [47], chemical distribution [15], scattering [12], polarization [84] and birefringence [16]. This integration can enable the learning of simultaneous information fusion of multiple informative properties to predict the desired digital stain in a single image. By combining label-free information of various scales, microscopic and mesoscopic structures could be revealed at the same time, which could be potentially beneficial to achieve higher

staining sensitivity and accuracy. The integration requires innovation from both the hardware system that is engineered to capture multi-modality images simultaneously as well as algorithms that optimize information fusion using advanced deep learning architectures.

Lastly, enhancing the quantitative analysis capabilities can enable more objective and reliable assessments of digital staining methods. This includes developing algorithms to quantify stain intensities, cellular features, and pathological features of specific stains. Different from traditional image quality metrics that are developed for general natural images, such quantification needs metrics that are meaningful to cell biologists, histopathologists and neuroscientists. To do so, establishing standardized validation procedures for digital staining is essential to ensure the quantitative assessment of results. Collaboration among computational imaging researchers, biologists and pathologists can help in defining guidelines and best practices for standard quantitative image analysis at both cellular level and tissue level. Furthermore, conducting clinical studies and validations to assess the diagnostic accuracy and clinical utility of digital staining is crucial for its widespread adoption.

Overall, the future of the digital staining field is promising and holds great potential for transforming high-content imaging, drug discovery and digital pathology. From the cellular scale digital staining, researchers can assess the impact of the compounds on specific cellular markers, protein expression, or other relevant staining patterns, which allow researchers to identify markers associated with specific drug responses or disease phenotypes. This enables the rapid screening of large compound libraries, aiding in the identification of potential drug candidates. From the tissue scale digital staining, by providing digital access to a wide range of stains and markers, pathologists can gain more insights into tissue samples, identify subtle features, and make more accurate and informed diagnoses. This can lead to improved pa-

tient care and outcomes. Digital staining techniques can be further integrated with computational pathology methods, enabling automated analysis and interpretation of digital stains. Deep learning models can leverage the rich information provided by digital staining to assist pathologists in tasks such as lesion grading and identification of rare or subtle features. The continued advancements in label-free imaging technology, such as the development of multi-scale, multi-modality imaging systems, will further enhance the capabilities of digital staining. These improvements can lead to better sensitivity of cell/tissue properties and spatial morphology and therefore higher-quality digital stains. The future of digital staining is likely to witness a convergence of advancements in computational imaging technology, high-content screening, computational pathology, and deep learning, enabling accelerated drug discovery, more accurate diagnoses, and improved patient care. As the field progresses, it will be important to address technical challenges, validate the clinical utility, and promote collaborations among researchers, biologists, pathologists, and regulatory bodies to realize the full potential of digital staining.

References

- [1] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of signal processing*. Cambridge University Press, 2014.
- [2] C. A. Bouman, *Foundations of computational imaging: a model-based approach*. SIAM, 2022.
- [3] A. Bhandari, A. Kadambi, and R. Raskar, *Computational Imaging*. MIT Press, 2022.
- [4] G. Barbastathis, A. Ozcan, and G. Situ, “On the use of deep learning for computational imaging,” *Optica*, vol. 6, no. 8, pp. 921–943, 2019.
- [5] M. Kellman, K. Zhang, E. Markley, J. Tamir, E. Bostan, M. Lustig, and L. Waller, “Memory-efficient learning for large-scale computational imaging,” *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1403–1414, 2020.
- [6] Y. Li and L. Tian, “Computer-free computational imaging: optical computing for seeing through random media,” *Light: Science & Applications*, vol. 11, no. 1, p. 37, 2022.
- [7] Y. Li, Y. Xue, and L. Tian, “Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media,” *Optica*, vol. 5, no. 10, pp. 1181–1190, 2018.
- [8] Y. Li, S. Cheng, Y. Xue, and L. Tian, “Displacement-agnostic coherent imaging through scatter with an interpretable deep neural network,” *Optics Express*, vol. 29, no. 2, pp. 2244–2257, 2021.
- [9] Y. Li, G. N. McKay, N. J. Durr, and L. Tian, “Diffuser-based computational imaging funduscope,” *Optics Express*, vol. 28, no. 13, pp. 19 641–19 654, 2020.
- [10] Y. Xue, S. Cheng, Y. Li, and L. Tian, “Reliable deep-learning-based phase imaging with uncertainty quantification,” *Optica*, vol. 6, no. 5, pp. 618–629, 2019.
- [11] S. Chang, D. Varadarajan, J. Yang, I. A. Chen, S. Kura, C. Magnain, J. C. Augustinack, B. Fischl, D. N. Greve, D. A. Boas *et al.*, “Scalable mapping of myelin and neuron density in the human brain with micrometer resolution,” *Scientific Reports*, vol. 12, no. 1, p. 363, 2022.

- [12] S. Chang, J. Yang, A. Novoseltseva, X. Fu, C. Li, S.-C. Chen, J. C. Augustinack, C. Magnain, B. Fischl, A. C. McKee *et al.*, “Multi-scale label-free human brain imaging with integrated serial sectioning polarization sensitive optical coherence tomography and two-photon microscopy,” *bioRxiv*, pp. 2023–05, 2023.
- [13] J. Yang, I. A. Chen, S. Chang, J. Tang, B. Lee, K. Kılıç, S. Sunil, H. Wang, D. Varadarajan, C. Magnain *et al.*, “Improving the characterization of ex vivo human brain optical properties using high numerical aperture optical coherence tomography by spatially constraining the confocal parameters,” *Neurophotonics*, vol. 7, no. 4, pp. 045 005–045 005, 2020.
- [14] J. Yang, S. Chang, I. A. Chen, S. Kura, G. A. Rosen, N. A. Saltiel, B. R. Huber, D. Varadarajan, Y. Balbastre, C. Magnain *et al.*, “Volumetric characterization of microvasculature in ex vivo human brain samples by serial sectioning optical coherence tomography,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 12, pp. 3645–3656, 2022.
- [15] H. Lin, H. J. Lee, N. Tague, J.-B. Lugagne, C. Zong, F. Deng, J. Shin, L. Tian, W. Wong, M. J. Dunlop *et al.*, “Microsecond fingerprint stimulated raman spectroscopic imaging by ultrafast tuning and spatial-spectral learning,” *Nature communications*, vol. 12, no. 1, p. 3052, 2021.
- [16] N. Blanke, V. Go, D. L. Rosene, and I. J. Bigio, “Quantitative birefringence microscopy for imaging the structural integrity of cns myelin following circumscribed cortical injury in the rhesus monkey,” *Neurophotonics*, vol. 8, no. 1, pp. 015 010–015 010, 2021.
- [17] R. Liu, Y. Sun, J. Zhu, L. Tian, and U. S. Kamilov, “Recovery of continuous 3d refractive index maps from discrete intensity-only measurements using neural fields,” *Nature Machine Intelligence*, vol. 4, no. 9, pp. 781–791, 2022.
- [18] Y.-Z. Liu, F. A. South, Y. Xu, P. S. Carney, and S. A. Boppart, “Computational optical coherence tomography,” *Biomedical optics express*, vol. 8, no. 3, pp. 1549–1574, 2017.
- [19] 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, p. 57, 2023.
- [20] C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, “Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy,” *Nature methods*, vol. 15, no. 11, pp. 917–920, 2018.
- [21] E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O’neil, K. Shah, A. K. Lee *et al.*, “In silico labeling: predicting fluorescent labels in unlabeled images,” *Cell*, vol. 173, no. 3, pp. 792–803, 2018.

- [22] S. Cheng, S. Fu, Y. M. Kim, W. Song, Y. Li, Y. Xue, J. Yi, and L. Tian, “Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy,” *Science advances*, vol. 7, no. 3, p. eabe0431, 2021.
- [23] A. Somani, A. A. Sekh, I. S. Opstad, Å. B. Birgisdottir, T. Myrmel, B. S. Ahluwalia, A. Horsch, K. Agarwal, and D. K. Prasad, “Virtual labeling of mitochondria in living cells using correlative imaging and physics-guided deep learning,” *Biomedical Optics Express*, vol. 13, no. 10, pp. 5495–5516, 2022.
- [24] 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, no. 1, p. 23, 2019.
- [25] Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk *et al.*, “Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning,” *Nature biomedical engineering*, vol. 3, no. 6, pp. 466–477, 2019.
- [26] Y. Rivenson, K. de Haan, W. D. Wallace, and A. Ozcan, “Emerging advances to transform histopathology using virtual staining,” *BME Frontiers*, vol. 2020, 2020.
- [27] 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, p. 78, 2020.
- [28] M. Chen, L. Tian, and L. Waller, “3d differential phase contrast microscopy,” *Biomedical optics express*, vol. 7, no. 10, pp. 3940–3950, 2016.
- [29] N. Shibata, S. D. Findlay, Y. Kohno, H. Sawada, Y. Kondo, and Y. Ikuhara, “Differential phase-contrast microscopy at atomic resolution,” *Nature Physics*, vol. 8, no. 8, pp. 611–615, 2012.
- [30] L. Tian and L. Waller, “Quantitative differential phase contrast imaging in an led array microscope,” *Optics express*, vol. 23, no. 9, pp. 11 394–11 403, 2015.
- [31] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, “Optical coherence tomography-principles and applications,” *Reports on progress in physics*, vol. 66, no. 2, p. 239, 2003.
- [32] J. M. Schmitt, “Optical coherence tomography (oct): a review,” *IEEE Journal of selected topics in quantum electronics*, vol. 5, no. 4, pp. 1205–1215, 1999.
- [33] V. J. Srinivasan, S. Sakadžić, I. Gorczynska, S. Ruvinskaya, W. Wu, J. G. Fujimoto, and D. A. Boas, “Quantitative cerebral blood flow with optical coherence tomography,” *Optics express*, vol. 18, no. 3, pp. 2477–2494, 2010.

- [34] L. Tian, X. Li, K. Ramchandran, and L. Waller, “Multiplexed coded illumination for fourier ptychography with an led array microscope,” *Biomedical optics express*, vol. 5, no. 7, pp. 2376–2389, 2014.
- [35] J. Li, J. Garfinkel, X. Zhang, D. Wu, Y. Zhang, K. De Haan, H. Wang, T. Liu, B. Bai, Y. Rivenson *et al.*, “Biopsy-free in vivo virtual histology of skin using deep learning,” *Light: Science & Applications*, vol. 10, no. 1, p. 233, 2021.
- [36] Y. Winetraub, E. Yuan, I. Terem, C. Yu, W. Chan, H. Do, S. Shevidi, M. Mao, J. Yu, M. Hong *et al.*, “Oct2hist: Non-invasive virtual biopsy using optical coherence tomography,” *medRxiv*, pp. 2021–03, 2021.
- [37] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [39] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [41] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.
- [42] A. Rana, G. Yauney, A. Lowe, and P. Shah, “Computational histological staining and destaining of prostate core biopsy rgb images with generative adversarial neural networks,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 828–834.
- [43] M. T. Shaban, C. Baur, N. Navab, and S. Albarqouni, “Staining: Stain style transfer for digital histological images,” in *2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019)*. IEEE, 2019, pp. 953–956.
- [44] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

- [45] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, “Contrastive learning for unpaired image-to-image translation,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*. Springer, 2020, pp. 319–345.
- [46] R. Cao, S. D. Nelson, S. Davis, Y. Liang, Y. Luo, Y. Zhang, B. Crawford, and L. V. Wang, “Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy,” *Nature Biomedical Engineering*, vol. 7, no. 2, pp. 124–134, 2023.
- [47] X. Li, G. Zhang, H. Qiao, F. Bao, Y. Deng, J. Wu, Y. He, J. Yun, X. Lin, H. Xie *et al.*, “Unsupervised content-preserving transformation for optical microscopy,” *Light: Science & Applications*, vol. 10, no. 1, p. 44, 2021.
- [48] M. Combalia Escudero, J. Pérez Ankar, A. García Herrera, L. Alos, V. Vilaplana Besler, F. Marqués Acosta, S. Puig, and J. Malvehy, “Digitally stained confocal microscopy through deep learning,” in *International Conference on Medical Imaging with Deep Learning: 8-10 July 2019, London, United Kingdom: proceedings of Machine Learning Research*. Microtome Publishing, 2019, pp. 121–129.
- [49] 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.
- [50] K. B. Ozyoruk, S. Can, B. Darbaz, K. Başak, D. Demir, G. I. Gokceler, G. Serin, U. P. Hacisalihoglu, E. Kurtuluş, M. Y. Lu *et al.*, “A deep-learning model for transforming the style of tissue images from cryosectioned to formalin-fixed and paraffin-embedded,” *Nature Biomedical Engineering*, pp. 1–13, 2022.
- [51] N. N. Boustany, S. A. Boppart, and V. Backman, “Microscopic imaging and spectroscopy with scattered light,” *Annual review of biomedical engineering*, vol. 12, pp. 285–314, 2010.
- [52] R. Kasprowicz, R. Suman, and P. O’Toole, “Characterising live cell behaviour: Traditional label-free and quantitative phase imaging approaches,” *The international journal of biochemistry & cell biology*, vol. 84, pp. 89–95, 2017.
- [53] Y. Park, C. Depeursinge, and G. Popescu, “Quantitative phase imaging in biomedicine,” *Nature photonics*, vol. 12, no. 10, pp. 578–589, 2018.
- [54] S. Uttam, H. V. Pham, J. LaFace, B. Leibowitz, J. Yu, R. E. Brand, D. J. Hartman, and Y. Liu, “Early prediction of cancer progression by depth-resolved nanoscale mapping of nuclear architecture from unstained tissue specimensnanoscale nuclear architecture predicts cancer progression,” *Cancer research*, vol. 75, no. 22, pp. 4718–4727, 2015.

- [55] A. Wax and K. J. Chalut, “Nuclear morphology measurements with angle-resolved low coherence interferometry for application to cell biology and early cancer detection,” *Analytical Cellular Pathology*, vol. 34, no. 5, pp. 207–222, 2011.
- [56] L. Cherkezyan, I. Capoglu, H. Subramanian, J. Rogers, D. Damania, A. Taflove, and V. Backman, “Interferometric spectroscopy of scattered light can quantify the statistics of subdiffractive refractive-index fluctuations,” *Physical review letters*, vol. 111, no. 3, p. 033903, 2013.
- [57] W. Song, A. Matlock, S. Fu, X. Qin, H. Feng, C. V. Gabel, L. Tian, and J. Yi, “Led array reflectance microscopy for scattering-based multi-contrast imaging,” *Optics letters*, vol. 45, no. 7, pp. 1647–1650, 2020.
- [58] A. Matlock, A. Sentenac, P. C. Chaumet, J. Yi, and L. Tian, “Inverse scattering for reflection intensity phase microscopy,” *Biomedical optics express*, vol. 11, no. 2, pp. 911–926, 2020.
- [59] S.-M. Guo, L.-H. Yeh, J. Folkesson, I. E. Ivanov, A. P. Krishnan, M. G. Keefe, E. Hashemi, D. Shin, B. B. Chhun, N. H. Cho *et al.*, “Revealing architectural order with quantitative label-free imaging and deep learning,” *elife*, vol. 9, p. e55502, 2020.
- [60] D. Kusumoto and S. Yuasa, “The application of convolutional neural network to stem cell biology,” *Inflammation and regeneration*, vol. 39, no. 1, pp. 1–7, 2019.
- [61] D. Kusumoto, M. Lachmann, T. Kunihiro, S. Yuasa, Y. Kishino, M. Kimura, T. Katsuki, S. Itoh, T. Seki, and K. Fukuda, “Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells,” *Stem cell reports*, vol. 10, no. 6, pp. 1687–1695, 2018.
- [62] C. L. Chen, A. Mahjoubfar, L.-C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali, “Deep learning in label-free cell classification,” *Scientific reports*, vol. 6, no. 1, p. 21471, 2016.
- [63] T. Blasi, H. Hennig, H. D. Summers, F. J. Theis, J. Cerveira, J. O. Patterson, D. Davies, A. Filby, A. E. Carpenter, and P. Rees, “Label-free cell cycle analysis for high-throughput imaging flow cytometry,” *Nature communications*, vol. 7, no. 1, p. 10256, 2016.
- [64] P. Eulenberg, N. Köhler, T. Blasi, A. Filby, A. E. Carpenter, P. Rees, F. J. Theis, and F. A. Wolf, “Reconstructing cell cycle and disease progression using deep learning,” *Nature communications*, vol. 8, no. 1, p. 463, 2017.
- [65] J. K. Zhang, Y. R. He, N. Sobh, and G. Popescu, “Label-free colorectal cancer screening using deep learning and spatial light interference microscopy (slim),” *APL photonics*, vol. 5, no. 4, p. 040805, 2020.

- [66] H. Kobayashi, C. Lei, Y. Wu, A. Mao, Y. Jiang, B. Guo, Y. Ozeki, and K. Goda, “Label-free detection of cellular drug responses by high-throughput bright-field imaging and machine learning,” *Scientific reports*, vol. 7, no. 1, p. 12454, 2017.
- [67] F. Buggenthin, F. Buettner, P. S. Hoppe, M. Endeke, M. Kroiss, M. Strasser, M. Schwarzfischer, D. Loeffler, K. D. Kokkaliaris, O. Hilsenbeck *et al.*, “Prospective identification of hematopoietic lineage choice by deep learning,” *Nature methods*, vol. 14, no. 4, pp. 403–406, 2017.
- [68] Y. Jo, S. Park, J. Jung, J. Yoon, H. Joo, M.-h. Kim, S.-J. Kang, M. C. Choi, S. Y. Lee, and Y. Park, “Holographic deep learning for rapid optical screening of anthrax spores,” *Science advances*, vol. 3, no. 8, p. e1700606, 2017.
- [69] A. Zaritsky, A. R. Jamieson, E. S. Welf, A. Nevarez, J. Cillay, U. Eskiocak, B. L. Cantarel, and G. Danuser, “Interpretable deep learning of label-free live cell images uncovers functional hallmarks of highly-metastatic melanoma,” *BioRxiv*, pp. 2020–05, 2020.
- [70] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman *et al.*, “Opportunities and obstacles for deep learning in biology and medicine,” *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018.
- [71] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” *arXiv preprint arXiv:1312.6034*, 2013.
- [72] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void-learning denoising from single noisy images,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2129–2137.
- [73] S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, “Imaging through glass diffusers using densely connected convolutional networks,” *Optica*, vol. 5, no. 7, pp. 803–813, 2018.
- [74] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [75] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. A. Lindquist, J. Moffat *et al.*, “Cellprofiler: image analysis software for identifying and quantifying cell phenotypes,” *Genome biology*, vol. 7, pp. 1–11, 2006.
- [76] F. Zanella, J. B. Lorens, and W. Link, “High content screening: seeing is believing,” *Trends in biotechnology*, vol. 28, no. 5, pp. 237–245, 2010.

- [77] G. M. Gaietta, B. N. Giepmans, T. J. Deerinck, W. B. Smith, L. Ngan, J. Llopis, S. R. Adams, R. Y. Tsien, and M. H. Ellisman, “Golgi twins in late mitosis revealed by genetically encoded tags for live cell imaging and correlated electron microscopy,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 47, pp. 17777–17782, 2006.
- [78] J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus *et al.*, “Data-analysis strategies for image-based cell profiling,” *Nature methods*, vol. 14, no. 9, pp. 849–863, 2017.
- [79] H. Wang, C. Magnain, S. Sakadžić, B. Fischl, and D. A. Boas, “Characterizing the optical properties of human brain tissue with high numerical aperture optical coherence tomography,” *Biomedical optics express*, vol. 8, no. 12, pp. 5617–5636, 2017.
- [80] C. J. Liu, W. Ammon, R. J. Jones, J. Nolan, R. Wang, S. Chang, M. P. Frosch, A. Yendiki, D. A. Boas, C. Magnain *et al.*, “Refractive-index matching enhanced polarization sensitive optical coherence tomography quantification in human brain tissue,” *Biomedical Optics Express*, vol. 13, no. 1, pp. 358–372, 2022.
- [81] C. Magnain, J. C. Augustinack, L. Tirrell, M. Fogarty, M. P. Frosch, D. Boas, B. Fischl, and K. S. Rockland, “Colocalization of neurons in optical coherence microscopy and nissl-stained histology in brodmann’s area 32 and area 21,” *Brain Structure and Function*, vol. 224, pp. 351–362, 2019.
- [82] C. Magnain, J. C. Augustinack, M. Reuter, C. Wachinger, M. P. Frosch, T. Ragan, T. Akkin, V. J. Wedeen, D. A. Boas, and B. Fischl, “Blockface histology with optical coherence tomography: a comparison with nissl staining,” *NeuroImage*, vol. 84, pp. 524–533, 2014.
- [83] C. Magnain, J. C. Augustinack, E. Konukoglu, M. P. Frosch, S. Sakadžić, A. Varjabedian, N. Garcia, V. J. Wedeen, D. A. Boas, and B. Fischl, “Optical coherence tomography visualizes neurons in human entorhinal cortex,” *NeuroPhotonics*, vol. 2, no. 1, pp. 015004–015004, 2015.
- [84] H. Wang, C. Magnain, R. Wang, J. Dubb, A. Varjabedian, L. S. Tirrell, A. Stevens, J. C. Augustinack, E. Konukoglu, I. Aganj *et al.*, “as-psoct: Volumetric microscopic imaging of human brain architecture and connectivity,” *Neuroimage*, vol. 165, pp. 56–68, 2018.
- [85] C. J. Liu, K. E. Williams, H. T. Orr, and T. Akkin, “Visualizing and mapping the cerebellum with serial optical coherence scanner,” *NeuroPhotonics*, vol. 4, no. 1, pp. 011006–011006, 2017.

- [86] H. Wang, J. Zhu, and T. Akkin, “Serial optical coherence scanner for large-scale brain imaging at microscopic resolution,” *Neuroimage*, vol. 84, pp. 1007–1017, 2014.
- [87] A. Odgaard, K. Andersen, F. Melsen, and H. J. G. Gundersen, “A direct method for fast three-dimensional serial reconstruction,” *Journal of microscopy*, vol. 159, no. 3, pp. 335–342, 1990.
- [88] K. de Haan, Y. Zhang, J. E. Zuckerman, T. Liu, A. E. Sisk, M. F. Diaz, K.-Y. Jen, A. Nobori, S. Liou, S. Zhang *et al.*, “Deep learning-based transformation of h&e stained tissues into special stains,” *Nature communications*, vol. 12, no. 1, p. 4884, 2021.
- [89] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, “Domain-adversarial training of neural networks,” *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.
- [90] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, “Nih image to imagej: 25 years of image analysis,” *Nature methods*, vol. 9, no. 7, pp. 671–675, 2012.
- [91] J. Fan, T. M. Dawson, and V. L. Dawson, “Cell death mechanisms of neurodegeneration,” *Neurodegenerative Diseases: Pathology, Mechanisms, and Potential Therapeutic Targets*, pp. 403–425, 2017.
- [92] A. L. Pistorio, S. H. Hendry, and X. Wang, “A modified technique for high-resolution staining of myelin,” *Journal of neuroscience methods*, vol. 153, no. 1, pp. 135–146, 2006.
- [93] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, “Color transfer between images,” *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.
- [94] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Goullart, and T. Yu, “scikit-image: image processing in python,” *PeerJ*, vol. 2, p. e453, 2014.
- [95] N. Kuninaka, M. Kawaguchi, M. Ogawa, A. Sato, K. Arima, S. Murayama, and Y. Saito, “Simplification of the modified g allyas method,” *Neuropathology*, vol. 35, no. 1, pp. 10–15, 2015.
- [96] P. A. Yushkevich, B. B. Avants, L. Ng, M. Hawrylycz, P. D. Burstein, H. Zhang, and J. C. Gee, “3d mouse brain reconstruction from histology using a coarse-to-fine approach,” in *Biomedical Image Registration: Third International Workshop, WBIR 2006, Utrecht, The Netherlands, July 9-11, 2006. Proceedings 3*. Springer, 2006, pp. 230–237.

- [97] S. Osechinskiy and F. Kruggel, "Slice-to-volume nonrigid registration of histological sections to mr images of the human brain," *Anatomy Research International*, vol. 2011, 2011.
- [98] E. Min, S. Ban, J. Lee, A. Vavilin, S. Baek, S. Jung, Y. Ahn, K. Park, S. Shin, S. Han *et al.*, "Serial optical coherence microscopy for label-free volumetric histopathology," *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.
- [99] Y. Zhang, L. Kang, W. Yu, V. T. Tsang, and T. T. Wong, "Three-dimensional label-free histological imaging of whole organs by microtomy-assisted autofluorescence tomography," *Iscience*, vol. 25, no. 1, p. 103721, 2022.
- [100] Y. Sun, S. You, X. Du, A. Spaulding, Z. G. Liu, E. J. Chaney, D. R. Spillman Jr, M. Marjanovic, H. Tu, and S. A. Boppart, "Real-time three-dimensional histology-like imaging by label-free nonlinear optical microscopy," *Quantitative Imaging in Medicine and Surgery*, vol. 10, no. 11, p. 2177, 2020.
- [101] N. Palomero-Gallagher and K. Zilles, "Cortical layers: Cyto-, myelo-, receptor-and synaptic architecture in human cortical areas," *Neuroimage*, vol. 197, pp. 716–741, 2019.
- [102] R. Turner, "Myelin and modeling: Bootstrapping cortical microcircuits," *Frontiers in neural circuits*, vol. 13, p. 34, 2019.
- [103] I. Itzkan, L. Qiu, H. Fang, M. M. Zaman, E. Vitkin, I. C. Ghiran, S. Salahuddin, M. Modell, C. Andersson, L. M. Kimerer *et al.*, "Confocal light absorption and scattering spectroscopic microscopy monitors organelles in live cells with no exogenous labels," *Proceedings of the National Academy of Sciences*, vol. 104, no. 44, pp. 17 255–17 260, 2007.
- [104] V. R. Singh, Y. A. Yang, H. Yu, R. D. Kamm, Z. Yaqoob, and P. T. So, "Studying nucleic envelope and plasma membrane mechanics of eukaryotic cells using confocal reflectance interferometric microscopy," *Nature communications*, vol. 10, no. 1, p. 3652, 2019.
- [105] P. Wang, R. K. Bista, W. E. Khalbuss, W. Qiu, S. Uttam, K. Staton, L. Zhang, T. A. Brentnall, R. E. Brand, and Y. Liu, "Nanoscale nuclear architecture for cancer diagnosis beyond pathology via spatial-domain low-coherence quantitative phase microscopy," *Journal of biomedical optics*, vol. 15, no. 6, pp. 066 028–066 028, 2010.
- [106] R. Ling, W. Tahir, H.-Y. Lin, H. Lee, and L. Tian, "High-throughput intensity diffraction tomography with a computational microscope," *Biomedical optics express*, vol. 9, no. 5, pp. 2130–2141, 2018.

- [107] J. Yi, A. J. Radosevich, J. D. Rogers, S. C. Norris, İ. R. Çapoğlu, A. Taflove, and V. Backman, “Can oct be sensitive to nanoscale structural alterations in biological tissue?” *Optics express*, vol. 21, no. 7, pp. 9043–9059, 2013.
- [108] Z. Wu, B. B. Chhun, G. Schmunk, C. N. Kim, L.-H. Yeh, T. Nowakowski, J. Zou, and S. B. Mehta, “Dynamorph: learning morphodynamic states of human cells with live imaging and sc-rnaseq,” *bioRxiv*, 2020.
- [109] C. Bakal, J. Aach, G. Church, and N. Perrimon, “Quantitative morphological signatures define local signaling networks regulating cell morphology,” *science*, vol. 316, no. 5832, pp. 1753–1756, 2007.
- [110] M. H. Rohban, S. Singh, X. Wu, J. B. Berthet, M.-A. Bray, Y. Shrestha, X. Varelas, J. S. Boehm, and A. E. Carpenter, “Systematic morphological profiling of human gene and allele function via cell painting,” *Elife*, vol. 6, p. e24060, 2017.
- [111] D. Damania, H. K. Roy, D. Kunte, J. A. Hurteau, H. Subramanian, L. Cherkezyan, N. Krosnjar, M. Shah, and V. Backman, “Insights into the field carcinogenesis of ovarian cancer based on the nanocytology of endocervical and endometrial epithelial cells,” *International journal of cancer*, vol. 133, no. 5, pp. 1143–1152, 2013.
- [112] R. Hollandi, A. Szkalicity, T. Toth, E. Tasnadi, C. Molnar, B. Mathe, I. Grexa, J. Molnar, A. Balind, M. Gorbe *et al.*, “nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer,” *Cell Systems*, vol. 10, no. 5, pp. 453–458, 2020.
- [113] R. Liu, S. Cheng, L. Tian, and J. Yi, “Deep spectral learning for label-free optical imaging oximetry with uncertainty quantification,” *Light: Science & Applications*, vol. 8, no. 1, p. 102, 2019.
- [114] S. Sun, K. Han, D. Kong, C. You, and X. Xie, “Mirnf: Medical image registration via neural fields,” *arXiv preprint arXiv:2206.03111*, 2022.
- [115] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 172–189.

CURRICULUM VITAE

