MICROSCOPY IMAGE REGISTRATION, SYNTHESIS AND SEGMENTATION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Chichen Fu

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Edward J. Delp, Chair

School of Electrical and Computer Engineering

Dr. Paul Salama

School of Electrical and Computer Engineering

Dr. Mary L. Comer

School of Electrical and Computer Engineering

Dr. Fengqing M. Zhu

School of Electrical and Computer Engineering

Approved by:

Dr. Pedro Irazoqui

Head of the School Graduate Program

ACKNOWLEDGMENTS

First of all, I would like to thank my doctoral advisor Professor Edward J. Delp for offering me the opportunity to join his research lab, Video and Image Processing Laboratory (VIPER), and under his supervision. I am grateful to him for his guidance, support, advice, and criticism. I am especially thankful for his trust in me and his encouragement to me to challenge myself, to overcome obstacles and to explore in new dimensions.

I would like to thank Professor Paul Salama for his inspiration and involvement in microscopy image analysis. I appreciate all his invaluable time and efforts for helping me with my research and paper. I would like to thank Professor Fengqing Zhu for her insightful suggestions on research ideas and my future career. I would like to thank Professor Mary Comer for her advice, support and encouragement.

I would like to thank Professor Kenneth W. Dunn for sharing his knowledge in biology. His feedback helped me to have a new understanding of the goals of my project.

I would like to thank all of my microscopy project team members, Dr. Neeraj Gadgil, Mr. Soonam Lee, Mr. David J. Ho and Ms. Shuo Han. It is truly an honor to work in this team. I would like to thank David for being a great friend and co-worker. We have been working through so many challenge together. I would like to thank Soonam for being a friend and helping me with my paper. I would like to thank Shuo for being a friend and involving in my research. I could not count how many nights we have been working together. Those will be the precious memory in my life. I would like to thank again for their support, encouragement and heartful advices for my research and my personal life.

I would like to specially thank Dr. Neeraj Gadgil for being a great mentor at my first year of PhD. I would like to specially thank Dr. Khalid Tahboub for helping me with writing my first paper.

Studying and working in VIPER have been a great experience. I would like to thank all my talented colleagues: Mr. Shaobo Fang, Mr. Yuhao Chen, Mr. Daniel Mas, Mr. Javier Ribera, Mr. David Guera Cobo, Dr. Albert Parra, Ms. Qingchaung Chen, Ms. Ruiting Shao, Ms. Jeehyun Choe, Ms. Dahjung Chung, Ms. Blanca Delgado, Ms. Di Chen, Mr. Jiaju Yue, Mr. He Li, Dr. Joonsoo Kim and Mr. Sri Kalyan Yarlagadda. I would like to further thank Mr. Shaobo Fang who has been a great friend and a wonderful co-worker.

I would like to thank my parent for giving me advices, motivating me when I am confused, criticizing me when I am complacent and supporting me when I am depressed. I would like to thank them for also being great friends to me. I would like to thank my uncle and aunt for taking care of me since my first entry to the United States of America. Without their support, it is impossible for me to study at Purdue. I would like to thank all of my family members for their unconditional love and support.

I would like to thank my fiance Chang Liu for supporting me and accompanying with me through my undergraduate and my graduate life. She helps me become a mature person.

This work was partially supported by a George M. O'Brien Award from the National Institutes of Health NIH/NIDDK P30 DK079312 and the endowment of the Charles William Harrison Distinguished Professorship at Purdue University.

The images used in section 3 for microscopy image registration were provided by Dr. Martin Oberbarnscheidt of the University of Pittsburgh and the Thomas E. Starzl Transplantation Institute.

Image data used in section 4 for nuclei segmentation were provided by different groups. Data-I was provided by Malgorzata Kamocka of Indiana University and was collected at the Indiana Center for Biological Microscopy. Data-II and Data-III were provided by Tarek Ashkar of the Indiana University School of Medicine. Data-IV was provided by Kenneth W. Dunn of the Indiana University School of Medicine. We gratefully acknowledge their help and cooperation.

TABLE OF CONTENTS

[1] This is joint work with Dr. David J. Ho and Ms. Shuo Han.

Page

[2] This is joint work with Ms. Shuo Han and Mr. Soonam Lee

[3] This is joint work with Ms. Shuo Han, Mr. Soonam Lee and Dr. David J. Ho

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABSTRACT

Fu, Chichen Ph.D., Purdue University, May 2019. Microscopy Image Registration, Synthesis and Segmentation. Major Professor: Edward J. Delp.

Fluorescence microscopy has emerged as a powerful tool for studying cell biology because it enables the acquisition of 3D image volumes deeper into tissue and the imaging of complex subcellular structures. Fluorescence microscopy images are frequently distorted by motion resulting from animal respiration and heartbeat which complicates the quantitative analysis of biological structures needed to characterize the structure and constituency of tissue volumes. This thesis describes a two pronged approach to quantitative analysis consisting of non-rigid registration and deep convolutional neural network segmentation. The proposed image registration method is capable of correcting motion artifacts in three dimensional fluorescence microscopy images collected over time. In particular, our method uses 3D B-Spline based non-rigid registration using a coarse-to-fine strategy to register stacks of images collected at different time intervals and 4D rigid registration to register 3D volumes over time. The results show that the proposed method has the ability of correcting global motion artifacts of sample tissues in four dimensional space, thereby revealing the motility of individual cells in the tissue.

We describe in thesis nuclei segmentation methods using deep convolutional neural networks, data augmentation to generate training images of different shapes and contrasts, a refinement process combining segmentation results of horizontal, frontal, and sagittal planes in a volume, and a watershed technique to enumerate the nuclei. Our results indicate that compared to 3D ground truth data, our method can successfully segment and count 3D nuclei. Furthermore, a microscopy image synthesis method based on spatially constrained cycle-consistent adversarial networks is used to efficiently generate training data. A 3D modified U-Net network is trained with a combination of Dice loss and binary cross entropy metrics to achieve accurate nuclei segmentation. A multi-task U-Net is utilized to resolve overlapping nuclei. This method was found to achieve high accuracy object-based and voxel-based evaluations.

1. INTRODUCTION

1.1 Optical Microscopy Background

Microscopy is considered to be a significant tool for biomedical research. Electron microscopy uses electrons and electromagnetic waves as a source of illumination to observe a sample. Since the wavelength of an electron is much shorter than the wavelength of visible light, electron microscopy has higher resolving power when imaging smaller structures. However, the higher energy of electron microscopy can damage a living biological structure. Therefore, optical microscopy is preferred when observing living biological structures from living specimens. Optical microscopy or light microscopy uses visible light and a system of lenses to project a magnified image of a sample onto the retina of the eye or a imaging device. Fluorescence microscopy is a form of optical microscopy that uses fluorescence and phosphorescence for visualizing subcellular structures in living species.

The basic principle of fluorescence is illustrated using a Jablonski diagram [1, 2] shown in figure 1.1. Electrons of fluorescent molecules at the ground state are excited to excited singlet states by absorbing photons emitted by a light source. This can happen in femtoseconds (10^{-15} seconds). In the next few picoseconds (10^{-12} seconds), the excited fluorescent molecules transfer vibrational energy to heat energy through the process of Vibrational Relaxation. Consequently, the excited electron collapses to ground state in different ways:

- Fluorescence emission: Most of the excited electrons collapse to the ground state with the emission of a photon in nanoseconds (10^{-9} seconds).

- Phosphorescence Emission: Some of the excited electrons enter the triplet state that can make the molecule chemically active often leading to photobleaching. Photobleaching is the phenomenon of permanent loss of fluorescence due to photon-induced chemical damage and covalent modification. The excited electrons return to the ground state by the phenomenon of phosphorescence. Phosphorescence emission usually lasts from a fraction of a second to minutes.

It is to be noted, the energy of the emitted photon is not the same as the energy of the absorbed photon due to energy lost during vibrational relaxation. The energy of photon is given by:

$$E = \frac{hc}{\lambda} \tag{1.1}$$

where h is Planck's constant, c is the speed of light, and λ is the wavelength. Since the energy of a emitted photon is less than that of a absorbed one, the wavelength of the photon from light source is shorter than the wavelength of the emitted photon.

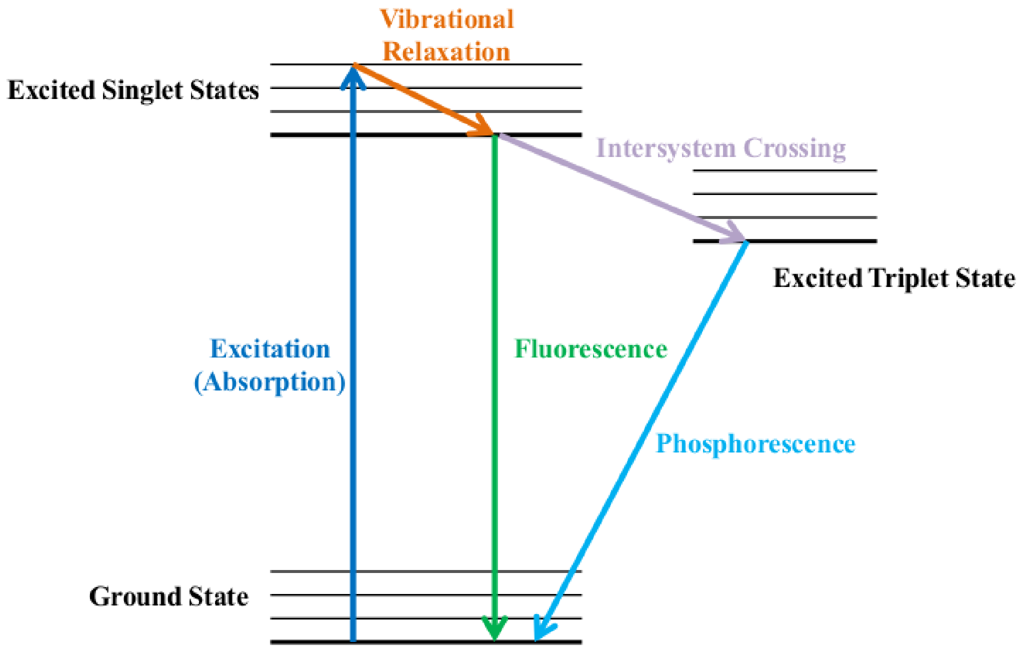


Fig. 1.1.: Jablonski diagram

Before using fluorescence microscopy to image a sample, the sample needs to be prepared to be fluorescent. One of the approaches is to use fluorophores, as known as fluroscent dyes or fluorescent molecules, to label different biological molecules of interest with different color. It enables imaging of different biological molecules of interest simultaneously. There are many fluorophores used in natural sciences since Richard Meyer created the term fluorophore in 1897. There are few criterias to the selection of flurophores. One is stokes shift. Stokes shift is defined as the wavelength difference between the maximum intensity of absorption and emission. Typically, The larger stokes shift of a flurophore, the easier that interference filter can isolate the light source and emitted light. The range of stokes shift of a flurophore can from a few to several hundred nanometers. Other important properties of fluorophores that need to be taken in consideration are extinction coefficient and quantum yield. The extinction coefficient measures the capacity for light absorption at a specific wavelength, while the quantum yield is the ratio of the number of fluorescence photons emitted to the number of photon absorbed. Higher quantum yield and extinction coefficient are desired for creating images with higher intensity from the identity light source. For example, Alexa and Cyanine dyes are popularly used because of their high quantum yield and high resistance to photobleaching. Also, some fluorophores are widely used because of their special characristics. For example, green fluorescent protein (GFP) [3–5], originally from jellyfish Aequorea victoria, *aequorea victoria*, and its mutated forms are widely used because it forms chromophore without additional cofactors or enzyme.

Biological specimens that are to be imaged are usually dyed with fluorphores and observed using a fluorescence microscope. Many different types of fluorescence microscopes exist such as the conventional widefield microscope, confocal microscope and Two-photon microscope. A conventional widefield microscope uses Kohler illumination [1] and magnifies the light emitted by the region of interest of the entire specimen. In conventional widefield microscope, secondary fluorescence emitted by the specimen often occurs through the excited volume and obscures resolution of features that lie in the objective focal plane. This problem can be worse when the sample is thicker than 2 micrometers. The out-of-focus light collected causes the lose of detail of imaging.

To solve this problem, Minsky invented a fluorescence microscope called ”confocal microscope” in 1955 [6]. A comparison of the operations of widefield microscope and confocal microscope is given in figure 1.2. A confocal microscope uses a pinholes to eliminate the out-of-focus light. This enables the confocal microscope to have higher signal contrast and preserves the details of specimen. However, the scanning speed of a confocal microscope is much slower. A new scanning method recently developed [7] that uses a spinning Nipkow disk to scans multiple points at a time is much more efficient than the traditional point-by-point scanning.

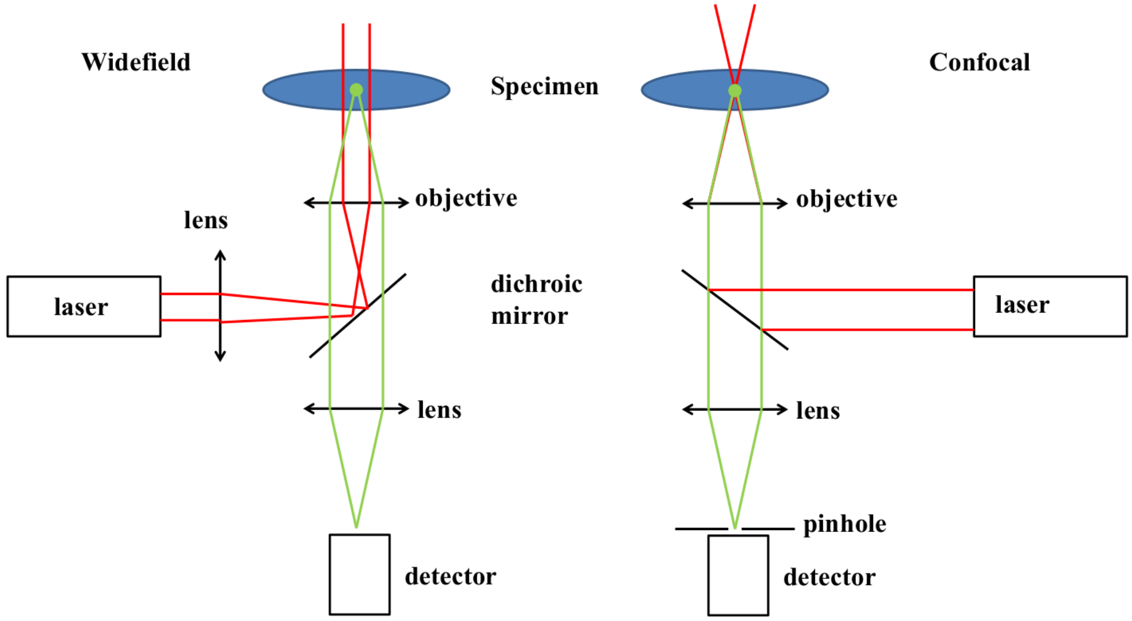


Fig. 1.2.: Comparison of the mechanism of widefield microscope and confocal microscope

The main drawback of a confocal microscope is the limitation of detecting photons at deep tissue depths. When the thickness of a sample is greater than the wavelength of excited light, the number of photons emitted deep in tissue could be reduced because of light scatting. Two-photon microscopy [1, 8–11] uses near-infrared light to excite fluorescence dyed samples to generate images in deeper tissue. The excitation of two photons of infrared light with longer wavelength aids in the significant reduction of the background signal generated by absorbing scattered fluorescence photons. Also, the use of two photon of infrared light enhances the penetration into the tissue. These two advantages of two-photon microscopy have enabled the visualization of sub-cellular structures located deeper in tissue. Moreover, in two-photon microscopy, two low energy photons excite a fluorescent molecule together. Unlike confocal microscopy, two-photon microscopy excitation is a non-linear process and the photon absorption rate is the square of the intensity of the light source. Therefore, excitation only occurs in a tiny focal volume which increases the contrast in 3D imaging [12]. Consequently, photobleaching and phototoxicity, two forms of photodamage are also reduced due to the small excitation region [1]. Multiphoton microscopy techniques have also been developed for visualizing deeper tissue [13].

1.2 Contributions of This Thesis

In this thesis, we discuss three different fields of microscopy image analysis including microscopy image registration, microscopy image synthesis and microscopy image segmentation. In Chapter 2, we describe a 4D microscopy image registration method. A microscopy image synthesis method using generative adversarial networks is discussed in Chapter 3. Three different microscopy image/volume segmentation methods are presented in Chapter 4. We also designed a microscopy image analysis system and it is described in Chapter 5.

The main contributions of this thesis are listed below:

- We describe a 4D registration method which can align image volumes in 3D + time space. In particular, we combine two registration approaches: a 3D non-rigid and a 4D rigid registration. The 3D non-rigid registration focuses on canceling motion artifacts between focal slices within a time volume and the 4D rigid registration focuses on canceling motion artifacts between time volumes. The results demonstrate that this method can correct motion artifacts generated by heart beating and respiration during data acquisition while preserving the original motion of biological structures.

- A Convolution Neural Network (CNN) based nuclei segmentation method is then described. This method is consisting of two stages, a CNN nuclei segmentation with majority voting refinement and watershed nuclei counting. Data augmentation is used to generate more training images with different shapes and contrast. The experimental results are validated using 3D manual annotated ground truth. We compared our method with two traditional methods, active contour and squassh. Our method achieves higher accuracy in terms of voxel evaluation.

- A spatial constrained cycle-consistent adversarial network (SpCycleGAN) is presented to synthesis nuclei volume. The SpCycleGAN generated realistic microscopy nuclei volumes without using paired training images. It also uses a spatial regulation to produce accurate nuclei location. The synthesized nuclei volume are used as training data for a 3D segmentation network. This 3D segmentation network was trained using a combination of binary cross entropy loss and a dice loss. The experimental results of segmentation nuclei show that our synthetic nuclei volumes are realistic and accurate.

- A Multi-task U-Net (MTU-Net) instance nuclei segmentation method is also described. First, our method generates binary nuclei segmentation and heatmap that contains the location information of the nuclei. Then, the heatmap was processed to be a binary location map of the nuclei. At last, a marker-controlled watershed uses binary location map of the nuclei as a marker to separate overlapped nuclei in the binary nuclei segmentation to produces an instance segmentation.

- A Distributed and Networked Analysis of Volumetric Image Data (DINAVID) system is developed to enable fast and accurate analysis of microscopy images. This system integrated with our segmentation methods and a 3D visualization tool. Also, it provides remote computing for biologist.

1.3 Publication Resulting From This Work

Journal Papers

1. **C. Fu**, S. Han, S. Lee, D. J. Ho, P. Salama, K. W. Dunn and E. J. Delp, "Three Dimensional Nuclei Synthesis and Instance Segmentation", *To be Submitted, IEEE Transactions on Medical Imaging.*

2. D. J. Ho, **C. Fu**, D. M. Montserrat, P. Salama and K. W. Dunn and E. J. Delp, "Sphere Estimation Network: Three Dimensional Nuclei Detection of Fluorescence Microscopy Images", *To be Submitted, IEEE Transactions on Medical Imaging.*

Conference Papers

1. **C. Fu**, N. Gadgil, K. K Tahboub, P. Salama, K. W. Dunn and E. J. Delp, "Four Dimensional Image Registration For Intravital Microscopy", *Proceedings of the Computer Vision for Microscopy Image Analysis workshop at Computer Vision and Pattern Recognition*, July 2016, Las Vegas, NV.

2. **C. Fu**, D. J. Ho, S. Han, P. Salama, K. W. Dunn, E. J. Delp, "Nuclei segmentation of fluorescence microscopy images using convolutional neural networks", *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 704-708, April 2017, Melbourne, Australia. DOI: 10.1109/ISBI.2017.7950617

3. **C. Fu**, S. Han, D. J. Ho, P. Salama, K. W. Dunn and E. J. Delp, "Three dimensional fluorescence microscopy image synthesis and segmentation", *Proceedings of the Computer Vision for Microscopy Image Analysis workshop at Computer Vision and Pattern Recognition*, June 2018, Salt Lake City, UT.

4. D. J. Ho, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Segmentation of Fluorescence Microscopy Images Using Three Dimensional Convolutional Neural Networks," *Proceedings of the Computer Vision for Microscopy*

Image Analysis (CVMI) workshop at Computer Vision and Pattern Recognition (CVPR), July 2017, Honolulu, HI. DOI: 10.1109/CVPRW.2017.116

5. D. J. Ho, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Detection and Segmentation of Fluorescence Microscopy Images Using Three Dimensional Convolutional Neural Networks", *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 418-422, April 2018, Washington, DC. DOI: 10.1109/ISBI.2018.8363606

6. S. Lee, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Tubule Segmentation of Fluorescence Microscopy Images Based on Convolutional Neural Networks with Inhomogeneity Correction," *Proceedings of the IS&T Conference on Computational Imaging XVI*, February 2018, Burlingame, CA.

7. D. J. Ho, S. Han, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Center-Extraction-Based Three Dimensional Nuclei Instance Segmentation of Fluorescence Microscopy Images," *Submitted To, Proceedings of the IEEE International Symposium on Biomedical Imaging*, April 2019, Venice, Italy.

8. S. Han, S. Lee, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Counting in Microscopy Images with Three Dimensional Generative Adversarial Networks, *To, Appear, Proceedings of the SPIE Conference on Medical Imaging*, February 2019, San Diego, California.

2. 4D IMAGE REGISTRATION

2.1 Related Work

Recent advances in fluorescence microscopy allow imaging biological processes as they occur in living animals [8,10,14]. Fluorescence microscopy has been particularly useful for studies of the immune system [15, 16]. An effective immune response depends upon the behavior of immune cells, whose actions result in a defensive response against pathogens such as bacteria or viruses. Intravital microscopy is uniquely capable of characterizing the migration, activity and interactions of immune cells, making it a powerful tool for understanding the immune function. Studies of immune cell motility typically involve acquiring images of a 3D volume of tissue collected over time. Cell tracking is then used to characterize and quantify the motility of fluorescently-labeled immune cells in the tissue volume. The ability to characterize cell motility within a volume of tissue in a living animal is frequently compromised by global movement of the tissue resulting from animal respiration and heartbeat. Global motion artifacts must be corrected before cell tracking using image registration [17].

For microscopy, image registration focuses on aligning images from different focal slices and images or volumes taken from different times. In general, the most frequently used registration techniques can be divided into two categories. Intensity-based registration and feature-based registration [17, 18].

Feature-based methods comprise the use of image features used for feature correspondence matching and the estimation of an affine transformation matrix that corresponds to the distortion [17, 18]. [19] uses robust scale invariant feature transform (SIFT) to extract features from input images and matches feature points with evidence accumulation. Moving Least Squares transformation was used in [19] to estimate the geometric parameters for transformation. The main difficulties of feature-based registration include choosing the features and matching them across the images. Specifically, for images that contain highly active live cells that are traveling in 3D space, feature selection and matching can be challenging. Feature-based methods have better performance when similar structures are present in the scene.

A point-based 3D registration method that cancels 3D global translations and rotation around the z-axis in microscopy images with live cells is described in [20]. It uses threshold-based features, a feature matching method described in [21], and least-squares estimation of the affine transformation. This method is computationally fast because it uses partial information within the images but fails when there are significant scene changes. In [22], 4D microscopy images are registered by (i) matching Z directional image slices at different time volumes to find Z direction translation and (ii) using 2D landmark-based feature matching to align temporal volumes in the X and Y directions.

Intensity-based registration methods are often associated with deformation models, affine transformations, search methods, and similarity metrics. Many rigid and non-rigid intensity-based registration methods have been developed using deformation models, search methods, and similarity metrics. Rigid registration is usually used to address and cancel global motion, such as global transitions and rotations. Non-rigid registration can be used to cancel global and also localized non-rigid body motions. Rigid registration is often used before non-rigid registration in order to address both localized and global motion artifacts. In [23], a non-rigid registration method that minimizes residual complexity is described. Many similarity metrics such as sum of squared difference (SSD), gradient differences (GD), gradient correlation (GC), pattern intensity (PI), and mutual information (MI) [24, 25] have been used. GD and GC are gradient based methods that work well on the images with significant gradient information. MI is an entropy related method that has been effectively used on MRI and PET images, PI requires high contrast of input images to achieve high performance [24], and SSD can work effectively under more relaxed constraints and with less computational cost.

Image registration can also be considered as a optimization problem of energy function over a set of geometric parameters. Different optimization strategies often give various computation time and final outcomes. [26,27] describes a registration method that uses quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer to minimize a energy function. In this method, rigid registration was first used to series of images followed by localized non-rigid registration technique that employs B-spline interpolation technique. BFGS optimization reduced computational complexity by estimating Hessian matrix instead of computing it directly. A voxel-based rigid registration method is described in [28] that uses modified Marquardt-Levenberg optimization with a coarse-to-fine strategy to register 2D images or 3D volumes. This method produces promising results with functional magnetic resonance imaging (fMRI) and intramodality positron emission tomography (PET) data, which are different from microscopy images in that fMRI and PET images usually contain well defined structures. [29] introduced a hybrid global-local optimization method to correct the motion artifacts for the whole brain images. This new optimization greatly reduce the possibility of misregistrations by preventing from being trapping in local minima.

Different from the methods described above, another rigid registration method for canceling motion artifacts of biological objects based on frequency domain techniques is described in [30]. Besides fMRI and PET image registration, many good works were done in microscopy image registrations. A non-rigid registration method that utilizes multi-channel temporal 2D and 3D microscopy images of cell nuclei to address global rigid and local non-rigid motion artifacts of cell nuclei is described in [31]. Yet another non-rigid registration method [32] that cancels motion artifacts of subcellular particles in live cell nuclei in temporal 2D and 3D microscopy images by using the extensions of an optic flow method. However, these techniques [30–32] are mainly used to register images that contain single cellular structure. An thin plate spline non-rigid registration method that registers images containing many live cells is described in [33], but it can only cancel the motion artifacts between successive z stack images.

One of the objectives of our work in general is to track live cells while preserving functional motion. In general, non-rigid registration techniques have the ability to correct local object motion, but may "over register" and distort biological functional motion. Rigid registration techniques alternatively can preserve the cell motion and also cancel global motion artifacts. The images used in this thesis consists of a time series of four-channel (spectral channels red, green, blue, and yellow) 3D fluorescence microscopy volumes of immune cells collected from a mouse kidney. To be clear, our dataset consists of 4 spectral channels, each spectral channel is a 3D volume and the 3D volumes for each spectral channel are acquired over time at regular time intervals. We have 61 time samples where each time sample consists of 4 spectral channels. For each spectral channel we have 11 focal slices in the z direction (depth) where each focal slice is 512 × 512 pixels. The focal slices are acquired serially. Three spectral channels of the dataset contain immune cells that are moving in 3D space over time and the other spectral channel contains relatively stable blood flow through a tubular shaped structure. The cells are highly active over time and motion artifacts can be observed. Since the biological functional motion of the cells is valuable, cell motion over time should be preserved after registration. As we describe below, we consider our registration problem as a combination of two registration problems, a 3D non-rigid registration and 4D rigid registration. The 3D non-rigid registration focuses on canceling motion artifacts between focal slices at different time volumes and the 4D rigid registration focuses on canceling motion artifacts between time volumes.

2.2 Proposed Method

Figure 2.1 shows the block diagram of our proposed method. Our method consists of 1D cubic convolution interpolation, 3D non-rigid registration, 3D Gaussian blur, adaptive histogram equalization, and 4D rigid registration.

We use the following notation to represent the images, $F^{z_q}_{t_n,b_m}$, where z_q, t_n, b_m represent focal slices (z dimension), time samples, and spectral channels, respectively,

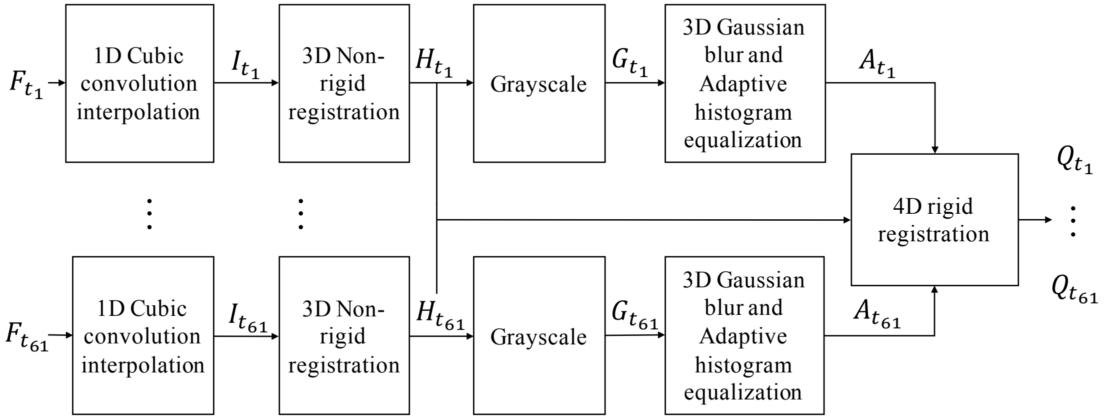

F_{t_1}
1D Cubic convolution interpolation
I_{t_1}
3D Non-rigid registration
H_{t_1}
Grayscale
G_{t_1}
3D Gaussian blur and Adaptive histogram equalization
A_{t_1}
4D rigid registration
Q_{t_1}
$Q_{t_{61}}$
$F_{t_{61}}$
1D Cubic convolution interpolation
$I_{t_{61}}$
3D Non-rigid registration
$H_{t_{61}}$
Grayscale
$G_{t_{61}}$
3D Gaussian blur and Adaptive histogram equalization
$A_{t_{61}}$

Fig. 2.1.: Block Diagram of the Proposed Method

where $q \in \{1, 2, \ldots, 11\}$, $n \in \{1, 2, \ldots, 61\}$, and $m \in \{1, 2, 3, 4\}$. The four channels 3D volume at the nth time sample in $F_{t_n,b_m}^{z_q}$ is denoted by F_{t_n}. For example, $F_{t_1,b_2}^{z_5}$ is the 512 × 512 pixel image representing the second spectral channel at the first time sample and the 5th focal slice within the volume. F_{t_1} contains 3D volumes from the four spectral channels collected at the first time sample with each volume consisting of 11 slices of 512 × 512 pixel images. Similarly, $I_{t_n,b_m}^{z_q}$, $H_{t_n,b_m}^{z_q}$, $Q_{t_n,b_m}^{z_q}$ denote the result of 1D cubic convolution interpolation, the result of 3D non-rigid registration, and the final 4D registration output, such that $q \in \{1, 2, \ldots, 41\}$, $n \in \{1, 2, \ldots, 61\}$, and $m \in \{1, 2, 3, 4\}$ (see Figure 2.1). The result of the 3D Gaussian blur and adaptive histogram equalization is denoted by $A_{t_n}^{z_q}$, such that $q \in \{1, 2, \ldots, 41\}$, and $n \in \{1, 2, \ldots, 61\}$. $A_{t_n}^{z_q}$ is a grayscale image. Please note that the total number of focal slices (as indicated by q) for $I_{t_n,b_m}^{z_q}$, $H_{t_n,b_m}^{z_q}$, $Q_{t_n,b_m}^{z_q}$ is 41 compared to 11 for $F_{t_n,b_m}^{z_q}$ since we use interpolation as described below. We mentioned above we register our images using 3D non-rigid registration and 4D rigid registration. F_{t_n} is up-sampled in the z direction to increase the resolution, the results is I_{t_n}. 3D non-rigid registration is then used to register z slices for each channel of a 3D volume at different time samples. The result is H_{t_n} is first transformed to grayscale images and then enhanced by using a 3D Gaussian blur and adaptive histogram equalization. 4D registration is used to estimate rigid body affine transformations for aligning A_{t_n}. The estimated affine transformations are then used to map H_{t_n} to the final result, Q_{t_n}, which are aligned in both time and the z direction.

2.2.1 Interpolation and 3D Non-Rigid Registration

To smooth our data, cubic convolution interpolation is used as a pre-processing step [34]. We up sample $F^{z_q}_{t_n,b_m}$ in the z direction by a factor of 4 to obtain $I^{z_q}_{t_n,b_m}$. Up sampling is done by inserting three data points between every two adjacent pixels in an image to produce 41 interpolated images for each spectral channel and time sample.

3D non-rigid registration is then used in the z direction to align the focal slices at each time sample. We use the non-rigid method described in our previous work [35, 36] because this method can effectively eliminate 3D non-rigid motion artifacts between focal slices. This technique initially starts with a rigid registration step and then uses localized non-rigid registration. The four spectral channels are transformed to grayscale images using the spectral channel weights described in [35, 36]. First, rigid registration is used to reduce global motion artifacts between images. This rigid registration uses Limited Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) Quasi-Newton optimization [37] to estimate the rigid body affine transformation. A localized non-rigid B-spline registration is then done on the results of the rigid registration described above to cancel the local non-rigid motion artifacts [35, 36].

In order to cancel the local non-rigid motion artifacts, images are deformed by establishing meshes of control points. A transformation is estimated to account for the movement of deformation fields using B-splines with L-BFGS Quasi-Newton optimization. A grid of control points is used in our method with 64 pixels spacings in X and Y directions.

2.2.2 Four Dimensional Rigid Registration

Having corrected motion artifacts between different focal planes, we use rigid registration to correct global translations and rotations. The input to this step is a set of multi-channel temporal 3D non-rigid registered volumes H_{t_n}. The multi-channel dataset used in this paper contained four channels: red, green, blue, and yellow. First, we transform the images in each time volume to composite grayscale images using a weighted sum:

$$G_{t_n} = \sum_{i=1}^{4} \frac{\bar{H}_{t_n,b_i}}{\sum_{j=1}^{4} \bar{H}_{t_n,b_j}} H_{t_n,b_i} \tag{2.1}$$

where $H_{t_n,b_i}, i \in \{1,2,3,4\}$ are the nth red channel, green channel, blue channel, and yellow channel 3D volumes, respectively. $\bar{H}_{t_n,b_i}, i \in \{1,2,3,4\}$ are the averaged pixel values of these channel volumes, respectively and G_{t_n} is the nth composite grayscale volume.

Biological structures are usually poorly defined in microscopy images. In order to create better defined structures, to improve registration performance, the grayscale images are 3D Gaussian blurred. Since our registration method is image intensity-based, low intensity and low contrast of the original images tend to cause the optimization method to be trapped in local minima, consequently producing incorrect affine transformation in intensity-based registration. To address this we enhance the grayscale images using adaptive histogram equalization (AHE) after the 3D Gaussian blur.

Affine transformations are then used between adjacent time volumes to minimize motion artifacts. The affine transformations are restricted to translations and rotations since we only focus on canceling rigid body motion artifacts at this stage. Denoting the translations and rotations in the X, Y, and Z directions, by $(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z)$

respectively, the corresponding translation and rotation matrices are given in Equations (2.2 - 2.6):

$$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta_x) & -\sin(\theta_x) & 0 \\ 0 & \sin(\theta_x) & \cos(\theta_x) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.2}$$

$$R_y = \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.3}$$

$$R_z = \begin{bmatrix} \cos(\theta_z) & \sin(\theta_z) & 0 & 0 \\ -\sin(\theta_z) & \cos(\theta_z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.4}$$

$$T = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.5}$$

$$M = R_x R_y R_z T \tag{2.6}$$

where R_x,R_y,R_z denote the rotation matrices around the X, Y, and Z axis respectively, T the translation matrix, and M the final affine transformation.

Broyden-Fletcher-Goldfarb-Shanno (BFGS) Quasi-Newton optimization was used to estimate the parameters m$(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z)$ by minimizing the sum of the squared differences (SSD) between different time volumes [38–41]. The optimal transformation is given by Equation (2.7):

$$\begin{aligned} M_n = \min_{M'} \sum_{x,y,z} [f(M', A_{t_n})(x, y, z) \\ -A_{t_{n-1}}(x, y, z)]^2 \end{aligned} \tag{2.7}$$

where A_{t_n} is the nth moving volume to be registered, $A_{t_{n-1}}$ the reference volume, $f(M^{'}, A_{t_n})$ the mapping that transforms current volume A_{t_n} by using transformation matrix $M^{'}$, and (x, y, z) are pixel coordinates.

When estimating the parameters of the affine transformation, we separate the process into two steps. First we estimate (t_x, t_y, θ_z) and $(\theta_x, \theta_y, t_z)$ separately with initial values of $(0, 0, 0)$. Second, we use the result from the previous step as an initial point of the final stage to obtain $(t_x, t_y, t_z, \theta_x, \theta_y, \theta_z)$. We have observed that using this strategy produces better results than doing the estimation in one step. Let M_i be the transformation estimated using the current time volume A_{t_i} and previous time volume $A_{t_{i-1}}$, and let T_n be the final affine transformation needed to correct motion artifacts between time volumes A_{t_1} and A_{t_n}. T_n is given by:

$$T_n = M_1 \times M_2 \times \cdots \times M_n \tag{2.8}$$

After the affine transformations of all the time volumes are estimated, the final registration outcomes are obtained by Equation 2.9:

$$Q_{t_n} = f(T_n, H_{t_n}) \tag{2.9}$$

where Q_{t_n} is the nth registered time volume. 3D cubic interpolation is used to transform pixels with non-integer coordinates in the function $f(\cdot)$. The registered volume's final size is the sum of the size of original volume and the maximum distance between original pixel locations and the transformed pixel coordinates in each direction.

2.2.3 3D Motion Vector Estimation - Validation

Validation of microscopy image registration can be daunting since ground-truth information is difficult to obtain on large image volumes. Block-matching is used to estimate motion vectors between the reference time volume and current time volume. This is somewhat similar to block matching techniques used in video compression. The current time volume and reference time volume are equally divided into blocks (sub-volumes). Each block in the current volume is matched with the corresponding adjacent blocks in the reference volume. Motion vectors are created to record the motion of corresponding blocks in the reference volume and the current volume that are matched. 3D time volumes are divided into sub-volumes with the size of i × j × k. A search window with the size of $(2p+1) \times (2p+1) \times p_z$ is created by setting the search range in the x,y, and z directions to (p, p, p_z).

To find the matching blocks and 3D motion vector $v = (i, j, k)$, the sum of absolute difference between reference block and current searching block is used:

$$\begin{aligned} v = \min_{a,b,c} \sum_{m,n,l} &|Q_{t_n}(x+m+a, y+n+b, z+l+c) \\ &- Q_{t_{n-1}}(x+m, y+n, z+l)| \end{aligned} \tag{2.10}$$

where $Q_{t_n}(x+m+a, y+n+b, z+l+c)$ is the current searching block in time volume Q_{t_n} centered at $(x+a, y+b, z+c)$, $Q_{t_{n-1}}(x+m, y+n, z+l)$ is the reference block centered at (x, y, z) in time volume $Q_{t_{n-1}}$, and (a, b, c) is the motion vector to be estimated. After the motion vectors are obtained, we create a 3D spherical histogram of the motion vectors with 36 × 36 bins to quantify the motion results.

2.3 Experimental Results

The images used in the experiments consists of a time series of four-channel (spectral channels red, green, blue, and yellow) 3D fluorescence microscopy volumes of immune cells collected from a mouse kidney. To be clear, our dataset consists of 4 spectral channels, each spectral channel is a 3D volume and the 3D volumes for each spectral channel are acquired over time at regular time intervals Samples of the four spectral channels are shown in Figure 2.2.

As shown in Figure 2.3, 1D cubic convolution interpolation is used to interpolate images in the z direction with the up-sampling factor of 4 in each 3D volume. Figure 2.3 (a) shows the YZ view of the green channel of an original 3D volume and the result of interpolation is shown in 2.3 (b). Note that the resulting 3D volume contains 4 times the number of images of the original.

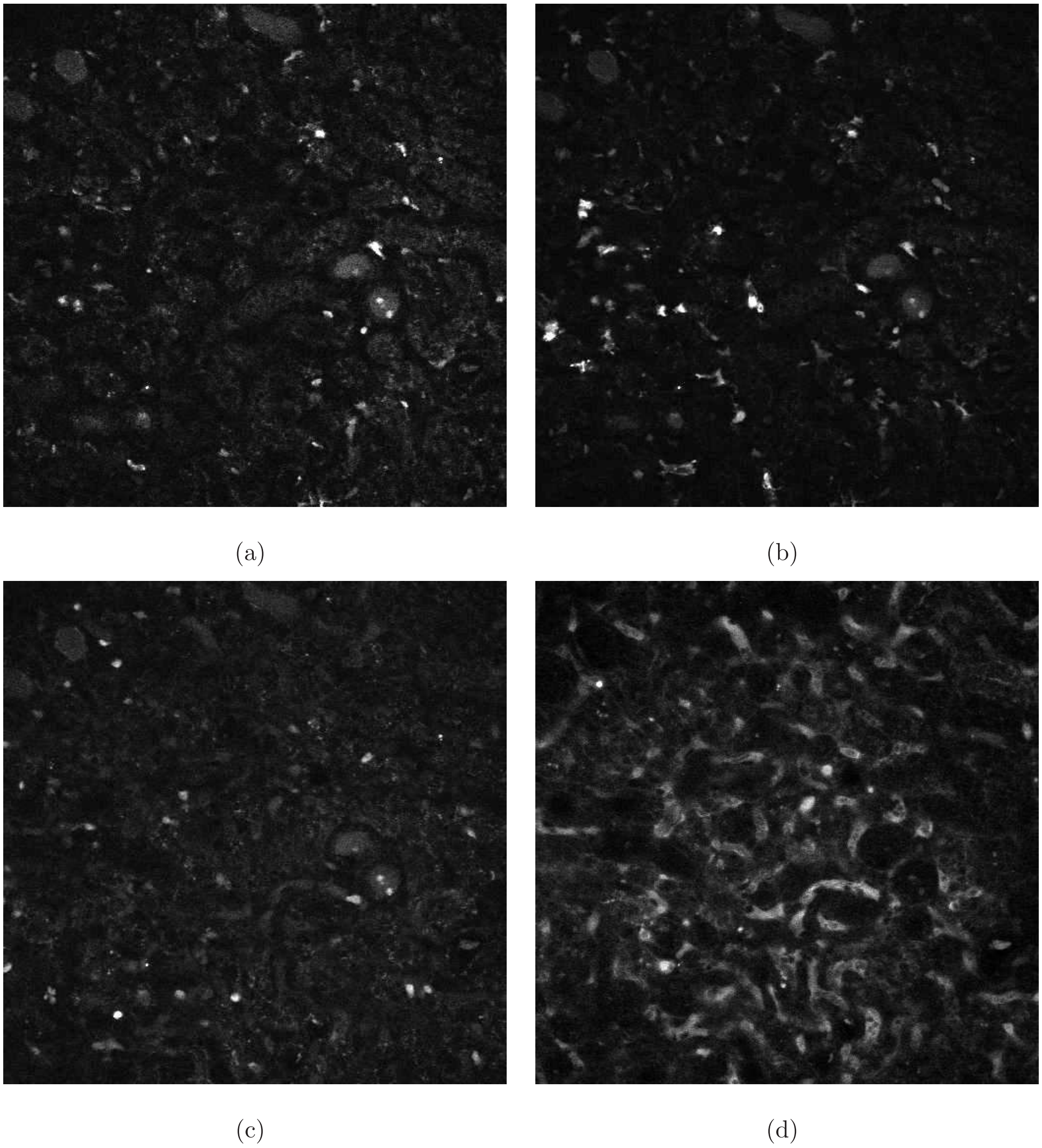
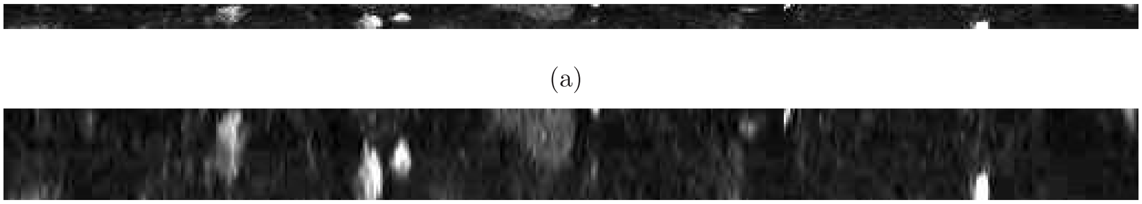


Fig. 2.2.: Grayscale versions of the four different spectral channels of the 6th focal slice of the 1st time volume of the original dataset. (a) Green channel, (b) Yellow channel, (c) Red channel, (d) Blue channel.

To evaluate the results of our registration method, we use maximum intensity projection (MIP) to project one 3D volume onto an image and many 3D volumes



(a)

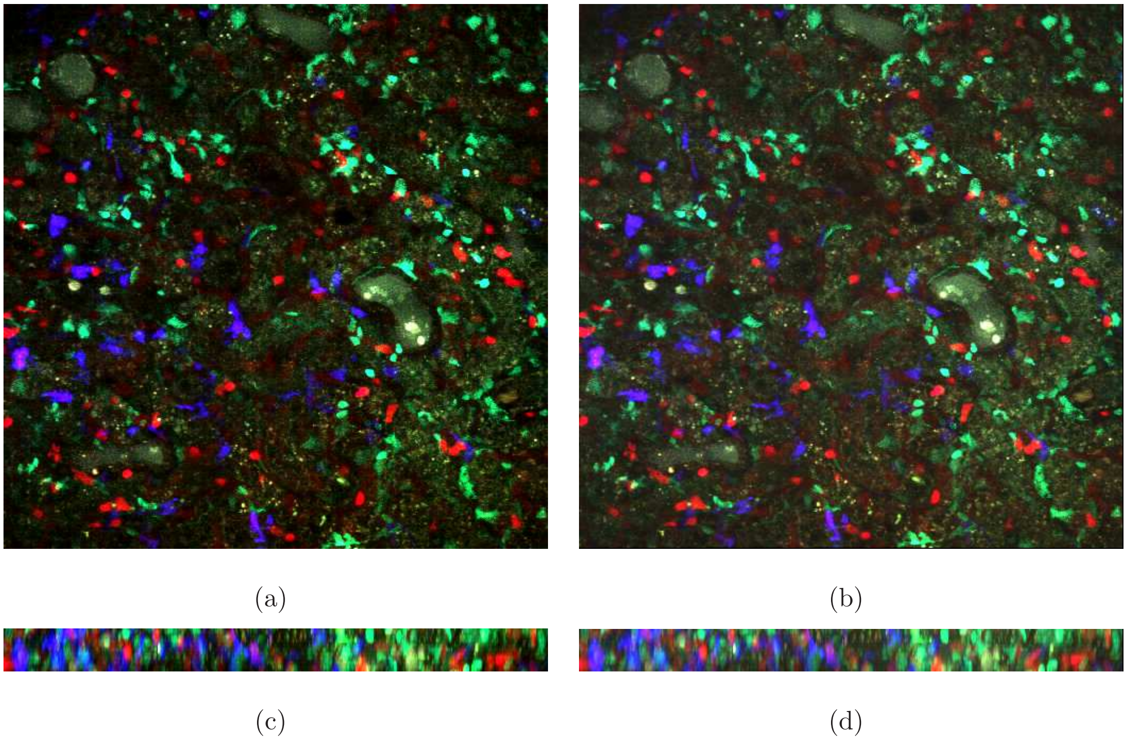
(b)

Fig. 2.3.: YZ view of the green channel of the original and the interpolated sample images. (a) Original, (b) Interpolated.

onto one 3D volume [35, 36]. The MIP of one 3D volume is obtained by selecting the maximum of all intensity values in one dimension (e.g. the z-direction or the time series) at each pixel location. The MIP of a 3D volume is used to show motion artifacts between focal slices, whereas the MIP of many 3D volumes representing different time samples is used to show motion artifacts between these volumes.

In Figure 2.4, we show the MIP of an original 3D volume and the MIP of the resulting 3D non-rigid registration. As we described in Section 2.2.1, 3D non-rigid registration is used to register focal slices within different 3D volumes. Since focal slices within different 3D volumes of our original dataset are well aligned initially, the impact of 3D non-rigid registration [35] can be observed in Figure 2.4. Temporal 3D microscopy data may not be well aligned in the z direction because all focal planes cannot be imaged at the same time instance. Therefore, 3D non-rigid registration is necessary to reduce motion artifacts between focal slices within different 3D volumes.

As shown in Figure 2.5 (a), the contrast of the four-channel composite sample image is very low and the biological structures are poorly defined. Therefore, a 3D Gaussian blur filter with $17 \times 17 \times 9$ rectangular window was used on the results of the 3D non-rigid registration followed by adaptive histogram equalization that employs $17 \times 17 \times 9$ rectangular window. Figure 2.5 (b) and (c) show the sample result of



(a)

(b)

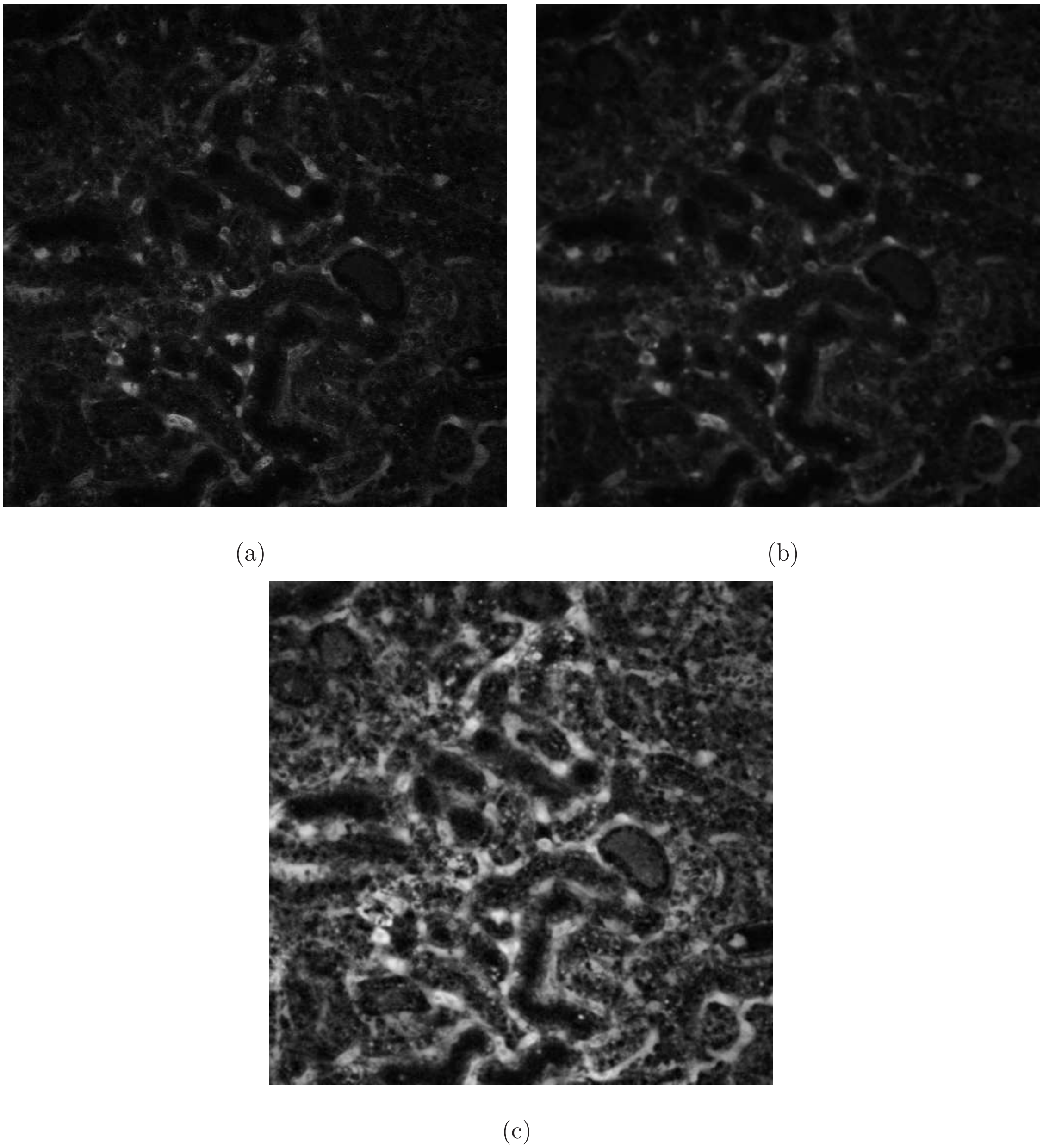
(c)

(d)

Fig. 2.4.: Sample images of our 3D non-rigid registration. (a) MIP of the sample original volume projected on XY plane, (b) MIP of the sample result of 3D non-rigid registration projected on XY plane, (c) MIP of the sample original volume projected on YZ plane, (d) MIP of the sample result of 3D non-rigid registration projected on YZ plane.

Gaussian blur and the sample result of adaptive histogram equalization. It can be observed that the sample image is enhanced.

Figure 2.6 (a) shows the MIPs projected on the XY plane of the original volumes at various samples. Figure 2.6 (c) shows the MIPs projected on the YZ plane. In Figure 2.6 (b) and (d), the MIPs of the results of our proposed 4D rigid registration method are shown. We also obtain the MIP of the entire 61 time volumes and use the ImageJ 3D viewer [42] to visualize it. Note that, this MIP is obtained by projecting a 4D volume on a 3D volume, whereas each of the MIPs shown in Figure 2.6 is obtained



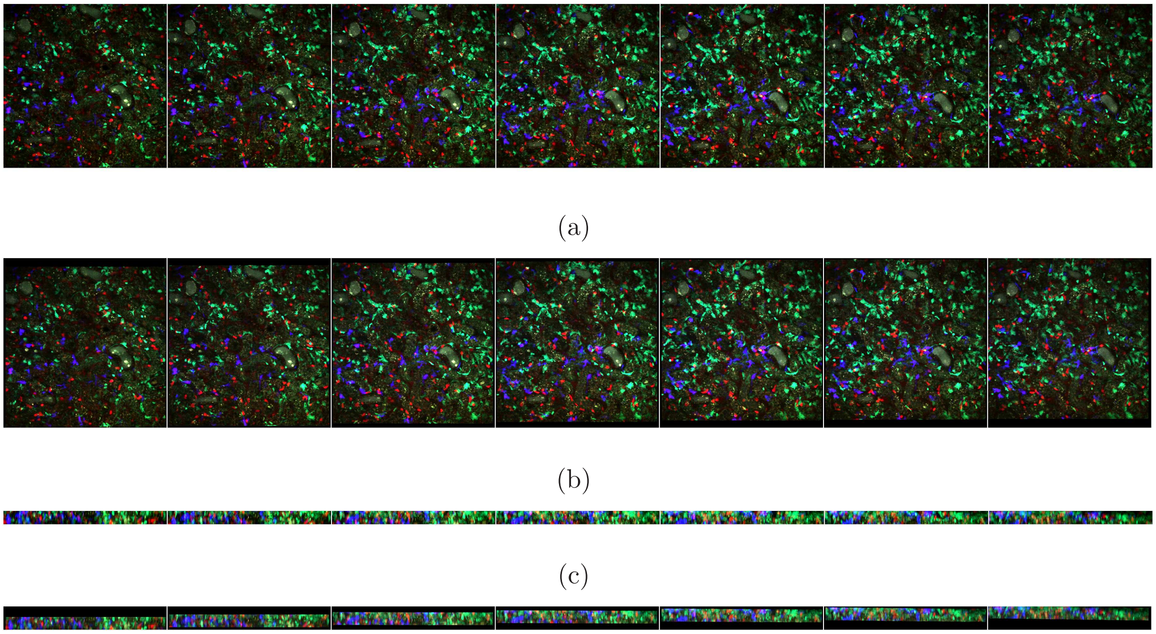
(a)

(b)

(c)

Fig. 2.5.: Sample results of pre-processing methods. (a) Composite grayscale original image, (b) 3D Gaussian blur, (c) Adaptive histogram equalization.

by projecting a 3D volume on a 2D image. The XY and YZ views of the MIP of the original time volumes are shown in Figure 2.7 (a) and (c) respectively. Figure 2.7 (b) and (d) show respectively the XY and YZ views of the MIP of the results of



(a)

(b)

(c)

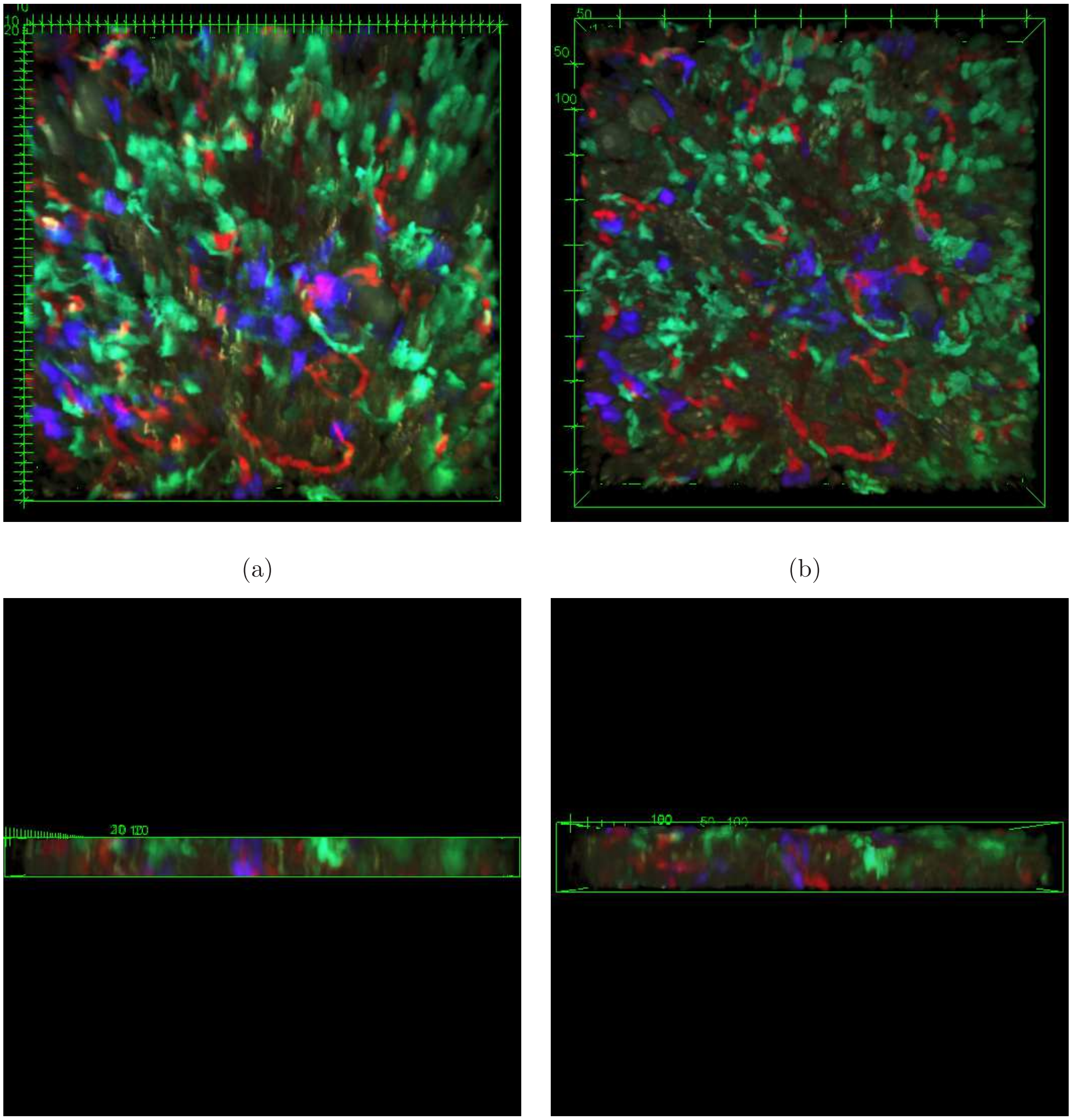
(d)

Fig. 2.6.: MIPs of the original time volumes and registered time volumes at time sample 1,11,21,31,41,51, and 61. (a) MIP of the original volumes projected on XY plane, (b) MIP of the result of 4D rigid registered volumes projected on XY plane, (c) MIP of the original volumes projected on YZ plane, (d) MIP of the result of 4D rigid registered volumes projected on YZ plane.

our proposed 4D rigid registration method. The MIP of original volumes appear to be smeared due to the global translations and rotations in time series. The MIP of registered volumes is sharper. The motility of cells can be observed in this MIP since it is a projection of the moving cells from different 3D volumes onto one volume. Note that the motions of cells are preserved during the registration process.

As shown in Figure 2.6 and Figure 2.7, our method successfully addressed the motion artifacts in our dataset and effectively cancel the motion artifacts in 4D space.

The average sum of squared differences (SSD) per pixel of the original and registered volumes are shown in Table **??**. The percentage improvement of the registered



(a)

(b)

(c)

(d)

Fig. 2.7.: Views of MIP volumes (using ImageJ 3D viewer). (a) XY view of original MIP volume, (b) XY view of 4D rigid registered MIP volume, (c) YZ view of original MIP volume, (d) YZ view of 4D rigid registered MIP volume.

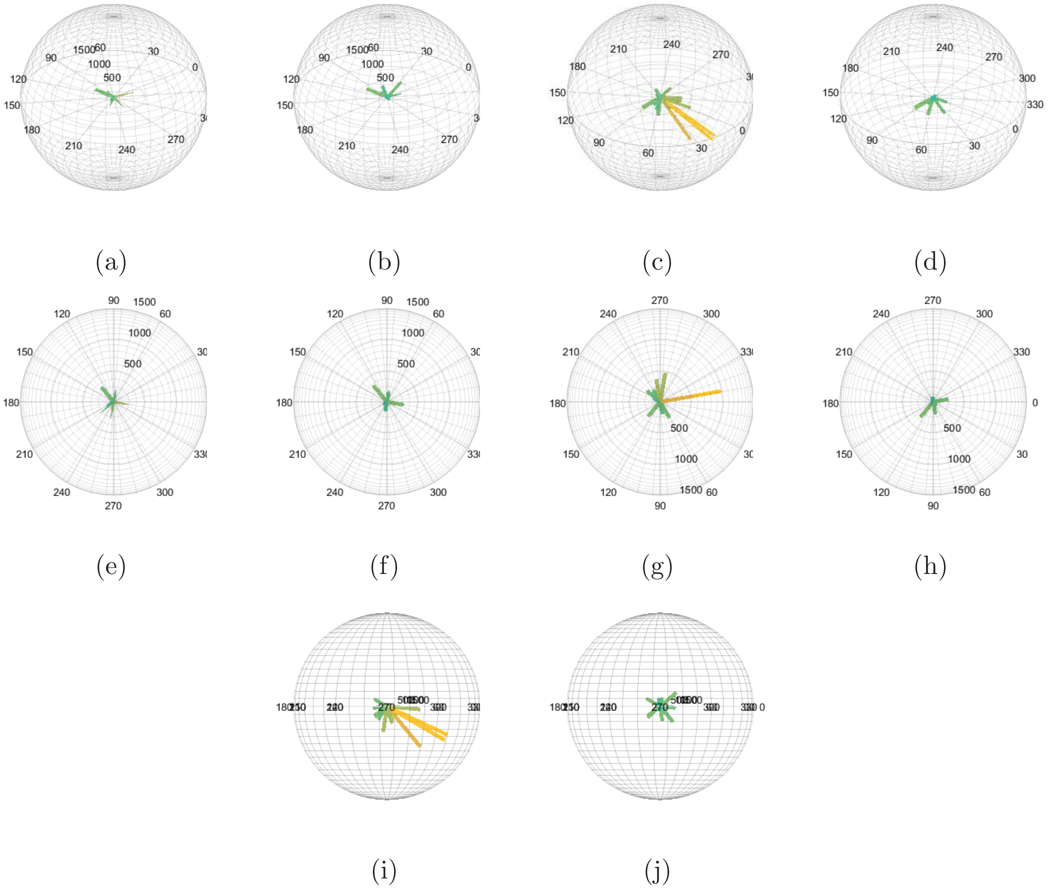
volumes as compared to the original is also shown. It can be observed that the aver-

Table 2.1.: Average SSD per pixel of different sample time volumes before and after registration and percentage of improvement.

Time point #	Average SSD per pixel before registration	Average SSD per pixel after registration	Improvement percentage (%)
11	7.88	6.59	16.41
21	9.45	8.71	7.84
31	10.29	8.54	16.94
41	12.52	8.79	29.76
51	9.36	7.98	14.79
61	8.76	7.08	19.12

age SSD per pixel decreases after 4D rigid registration indicating that the similarity between the reference and moving volumes is increased.

In addition, three dimensional motion vector analysis is used to validate the registration results as described in Section 2.2.3. Three dimensional motion vectors are obtained between adjacent time volumes using (16, 16, 8) as block size and $(4, 4, 4)$ as search window. Three dimensional spherical histograms are shown in Figures 2.8 and 2.9 using 36 × 36 bins of directions, each bin has range of 10 degrees. Various views of the three dimensional spherical histograms are shown in Figures 2.8 and 2.9 to help visualize the results. We observe that estimated motion are significantly reduced.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j)

Fig. 2.8.: 3D spherical histograms of motion vectors using time volume 9 as the moving volume and time volume 8 as the reference volume. (a) histogram of original volume in the view from top, (b) histogram of registered volume in the view from top, (c) histogram of original volume in the view from bottom, (d) histogram of registered volume in the view from bottom, (e) histogram of original volume in +XY view, (f) histogram of registered volume in +XY view, (g) histogram of original volume in -XY view, (h) histogram of registered volume in -XY view, (i) histogram of original volume in XZ view, (j) histogram of registered volume in XZ view.

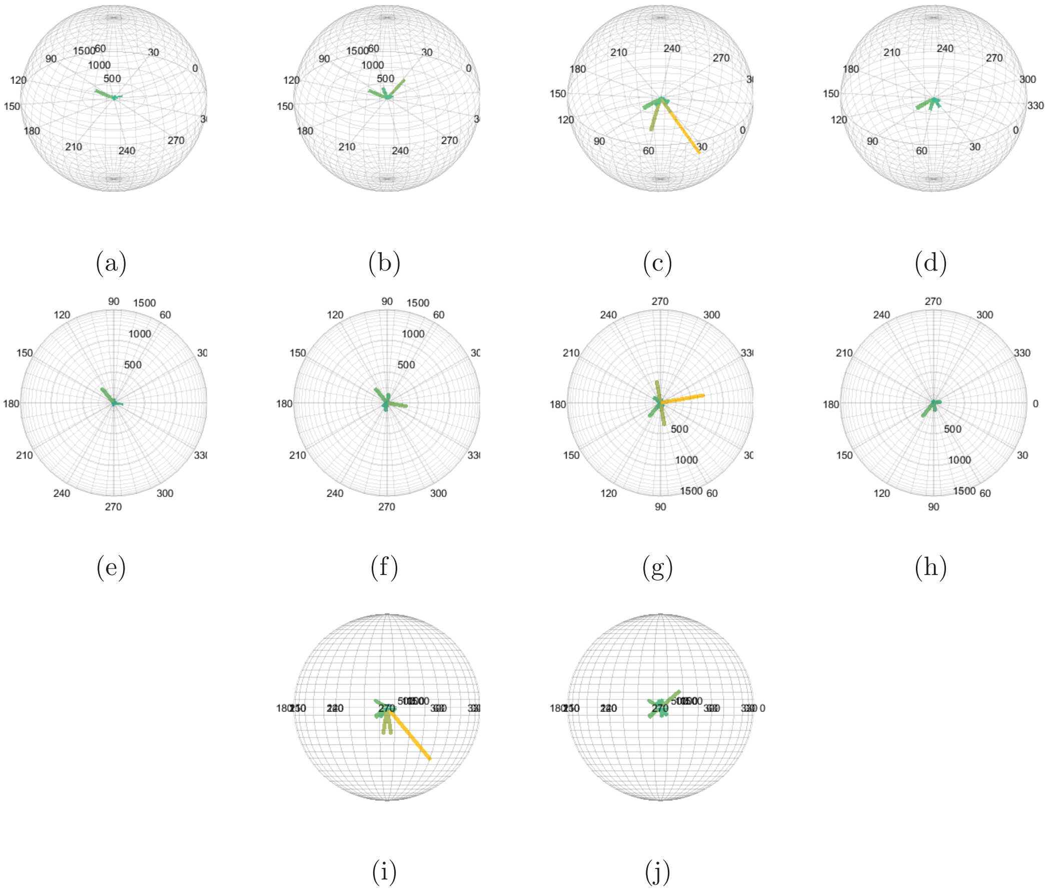


Fig. 2.9.: 3D spherical histograms of motion vectors using time volume 30 as the moving volume and time volume 29 as the reference volume. (a) histogram of original volume in the view from top, (b) histogram of registered volume in the view from top, (c) histogram of original volume in the view from bottom, (d) histogram of registered volume in the view from bottom, (e) histogram of original volume in +XY view, (f) histogram of registered volume in +XY view, (g) histogram of original volume in -XY view, (h) histogram of registered volume in -XY view, (i) histogram of original volume in XZ view, (j) histogram of registered volume in XZ view.

3. NUCLEI VOLUME SYNTHESIS

3.1 Related Work

Machine learning based methods require gigantic amounts of annotated data for training. However, annotation of microscopy images for segmentation is time consuming and requires expertise in biology. To address this problem, [43] generated synthetic microscopy volume by modeling electronic noise and blurring artifacts. However, the synthetic data generated by this method cannot reflect the characteristics of the real fluorescence microscopy volume. Moreover, in [44] the morphology of microscopy images are artificially created using a generative model. In addition, synthetic time-lapse 2D nuclei images with motion modeling was presented in [45]. Yet, generating realistic synthetic microscopy image volumes poses a challenging problem because of various noises and irregular shape of biological structures presented in microscopy volumes.

More recently, generative adversarial network (GAN) [46] was introduced to generate set of images from random noise using two adversarial networks. The generator learns a mapping G from distribution $p_z(z)$ to the distribution of real data $p_{data}(x)$. The discriminator D learns a discriminative function that tells whether generated distribution $p_z(z)$ is real or not. G and D are trained with a minimax strategy of a value function $V(G, D)$.

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)}[logD(x)] + E_{z \sim p_z(z)}[1 - logD(G(z))]$$

D was trained to maximize $V(G, D)$ while G was trained to minimize function $V(G, D)$. Figure 3.1 shows the structure of GAN. Generative adversarial networks are becoming the most popular model for image synthesis with different applications. A deep convolutional GAN (DCGAN) was demonstrated in [47] for unsupervised

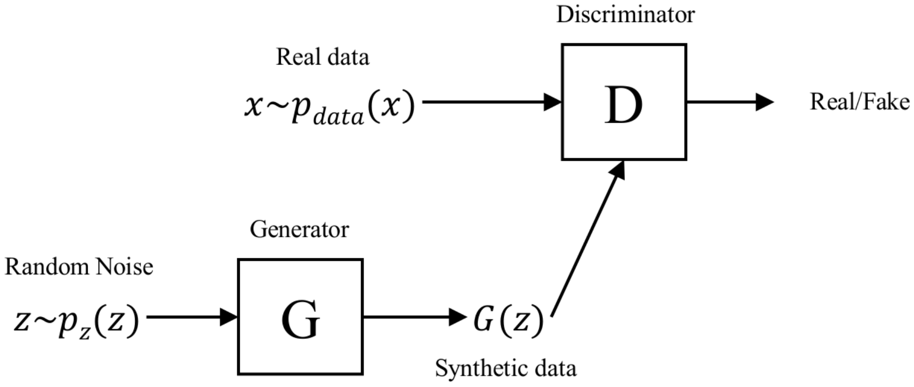


Fig. 3.1.: Structure of GAN

learning that constrained architectural topology to stabilize GAN training. Further, Wasserstein GAN (W-GAN) was described in [48] where utilizes Earth Mover distance instead of the probability distances and divergences to improve GAN training. In [49] both DCGAN and W-GAN were incorporated to synthesize cells from biological images. Later, Pix2Pix which introduces conditional GAN to improve the synthetic image quality using prior information of images was presented in [50]. One of the drawback of Pix2Pix [50] is that it requires paired training data. A cycle-consistent adversarial network (CycleGAN) [51] introduced a cycle consistent loss term to learn transformation using unpaired image training. Alternatively, a method that uses CycleGAN to translate CT segmentation to MRI segmentation was described in [52]. Additionally, [53] presented SpCycleGAN which adds spatial loss term to regularize the location of synthetic nuclei to improve nuclei segmentation results. Note that SpCycleGAN does not require any manually annotated groundtruth.

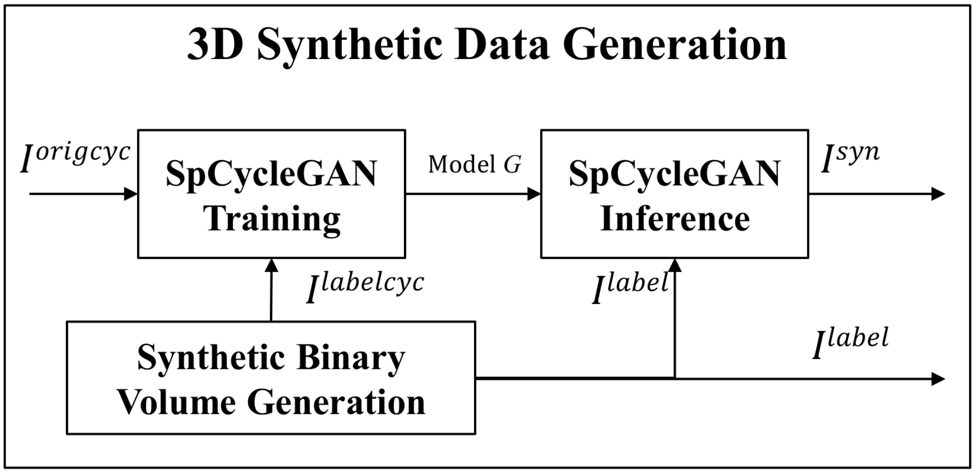


Fig. 3.2.: Block diagram of the proposed approach

3.2 Proposed Method

In order to train a segmentation network, a large amount of segmentation labels for each nuclei are required. However, this process is extremely laborious especially for labeling 3D volumes. In our method, we use a synthetic nuclei volume generation to create synthetic training volumes. The synthetic nuclei volume generation transfers from randomly generated binary nuclei volumes to real like microscopy volumes so that the binary nuclei volumes are accurately served as corresponding groundtruths for the nuclei in the generated real like microscopy volumes. We assume the shape of nucleus is an ellipsoid according to the observation of real microscopy volumes. The entire process consists of synthetic binary volume generation and synthetic microscopy volume generation. As shown in Figure 3.2, we first generate synthetic binary volumes, $I^{labelcyc}$, and synthetic heatmap volume, $I^{heatlabel}$ and then use them with a subvolume of the original image volumes, $I^{origcyc}$, to train a spatially constrained CycleGAN (SpCycleGAN) and obtain a generative model denoted as model G. This model G is used with another set of synthetic binary volume, I^{label}, to generate corresponding synthetic 3D volumes, I^{syn}. In this process, I^{syn}, I^{label} and $I^{heatlabel}$ are generated.

3.2.1 Synthetic Binary Volume Generation

Synthetic binary volume generation generates I^{label} by adding ellipsoid structures to a binary volume with random rotation and random translation [43]. Each of the ellipsoid structure represents single nucleus in the volume. The size of each ellipsoid structures is also randomly created within a preset range that obtained according to the observation of the characteristics of nuclei in I^{orig}. In addition, different nuclei are overlapped less than 5 voxels. Note that $I^{labelcyc}$ are generated similarly. While generating I^{label}, we also generate $I^{heatlabel}$. Here, $I^{heatlabel}$ is generated by the equation as below:

$$Ellip_n = \frac{x_n^2}{a_n^2} + \frac{y_n^2}{b_n^2} + \frac{z_n^2}{c_n^2}, \tag{3.1}$$

where x_n, y_n, and z_n are the sets of voxels on nth generated nuclei. a_n, b_n, and c_n are the semi-axis of the nth generated nuclei. We normalize $Ellip_n$ to be 255 at the center of the nuclei and then the value of $Ellip_n$ gradually decrease as the voxel getting close to the boundary of the nuclei. This normalization process can be performed by:

$$Ellip_n^{norm} = \frac{Ellip_n}{max\{\frac{1}{Ellip_n}\}} \times 255. \tag{3.2}$$

Finally, $I^{heatlabel}$ is obtained by adding each $Ellip_n^{norm}$ where $n \in \{1, \ldots, N\}$ where N is the total number of nuclei in the synthetic binary volume.

3.2.2 Synthetic Microscopy Volume Generation

In synthetic microscopy volume generation, we used both CycleGAN and our proposed SpCycleGAN.

CycleGAN

The CycleGAN is trained to generate a synthetic microscopy volume. CycleGAN uses a combination of discriminative networks and generative networks to solve a minimax problem by adding cycle consistency loss to the original GAN loss function as [46, 51]:

$$\begin{aligned}\mathcal{L}(G, F, D_1, D_2) = &\mathcal{L}_{\text{GAN}}(G, D_1, I^{labelcyc}, I^{origcyc}) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_2, I^{origcyc}, I^{labelcyc}) \\ &+ \lambda\mathcal{L}_{cyc}(G, F, I^{origcyc}, I^{labelcyc})\end{aligned} \tag{3.3}$$

where

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D_1, I^{labelcyc}, I^{origcyc}) = &\mathbb{E}_{I^{origcyc}}[\log(D_1(I^{origcyc}))] \\ &+ \mathbb{E}_{I^{labelcyc}}[\log(1 - D_1(G(I^{labelcyc})))]\end{aligned}$$

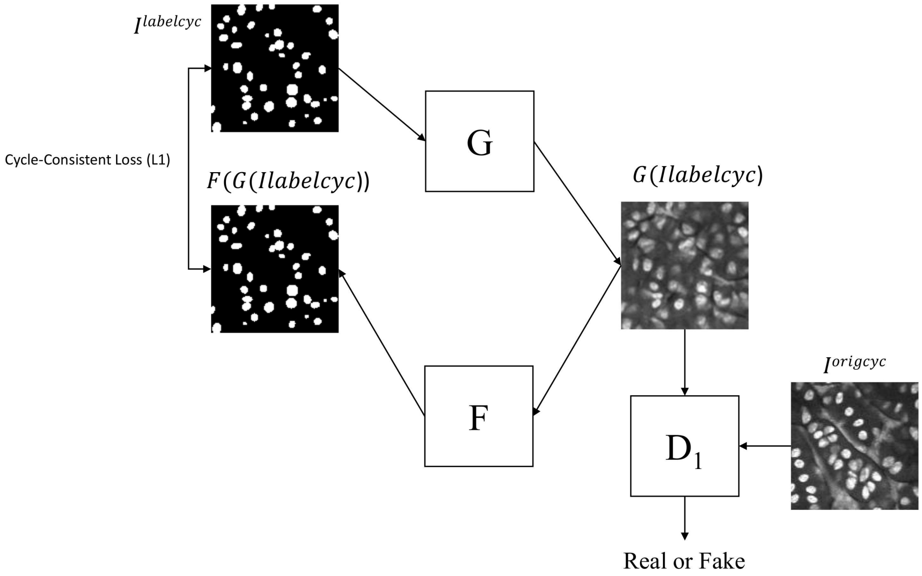


Fig. 3.3.: CycleGAN training path one

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(F, D_2, I^{origcyc}, I^{labelcyc}) &= \mathbb{E}_{I^{labelcyc}}[\log(D_2(I^{labelcyc}))] \\ &+ \mathbb{E}_{I^{origcyc}}[\log(1 - D_2(F(I^{origcyc})))]\end{aligned}$$

$$\begin{aligned}\mathcal{L}_{cyc}(G, F, I^{origcyc}, I^{labelcyc}) &= \mathbb{E}_{I^{labelcyc}}[||F(G(I^{labelcyc})) - I^{labelcyc}||_1] \\ &+ \mathbb{E}_{I^{origcyc}}[||G(F(I^{origcyc})) - I^{origcyc}||_1].\end{aligned}$$

Here, λ is a weight coefficient and $||\cdot||_1$ is L_1 norm. Note that Model G maps $I^{labelcyc}$ to $I^{origcyc}$ while Model F maps $I^{origcyc}$ to $I^{labelcyc}$. Also, D_1 distinguishes between $I^{origcyc}$ and $G(I^{labelcyc})$ while D_2 distinguishes between $I^{labelcyc}$ and $F(I^{origcyc})$. $G(I^{labelcyc})$ is an original like microscopy volume generated by model G and $F(I^{origcyc})$ is generated by model F that looks similar to a synthetic binary volume. Here, $I^{origcyc}$ and $I^{labelcyc}$ are unpaired set of images. In CycleGAN inference, I^{syn} is generated using the model

G on I^{label}. As previously indicated I^{syn} and I^{label} are a paired set of images. Here, I^{label} is served as a groundtruth volume corresponding to I^{syn}.

Spatially Constrained CycleGAN

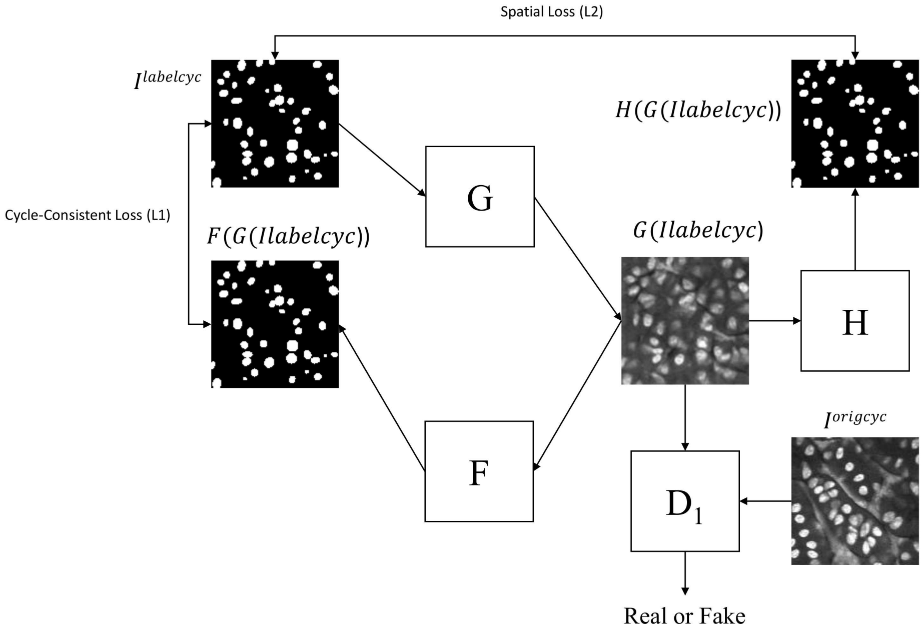


Fig. 3.4.: SpCycleGAN training path one

Figure 3.4 and 3.5 show the two training path of SpCycleGAN. CycleGAN and SpCycleGAN use the same training path two. Although the CycleGAN uses cycle consistency loss to constrain the similarity of the distribution of $I^{origcyc}$ and I^{syn}, CycleGAN does not provide enough spatial constraints on the locations of the nuclei. CycleGAN generates realistic synthetic microscopy images but a spatial shifting on the location of the nuclei in I^{syn} and I^{label} was observed. To create a spatial constraint on the location of the nuclei, a network H is added to the CycleGAN and takes $G(I^{labelcyc})$ as an input to generate a binary mask, $H(G(I^{labelcyc}))$. Here, the architecture of H is the same as the architecture of G. Network H minimizes a L_2 loss, $\mathcal{L}_{\text{Spatial}}$, between $H(G(I^{labelcyc}))$ and $I^{labelcyc}$. $\mathcal{L}_{\text{Spatial}}$ serves as a spatial regula-

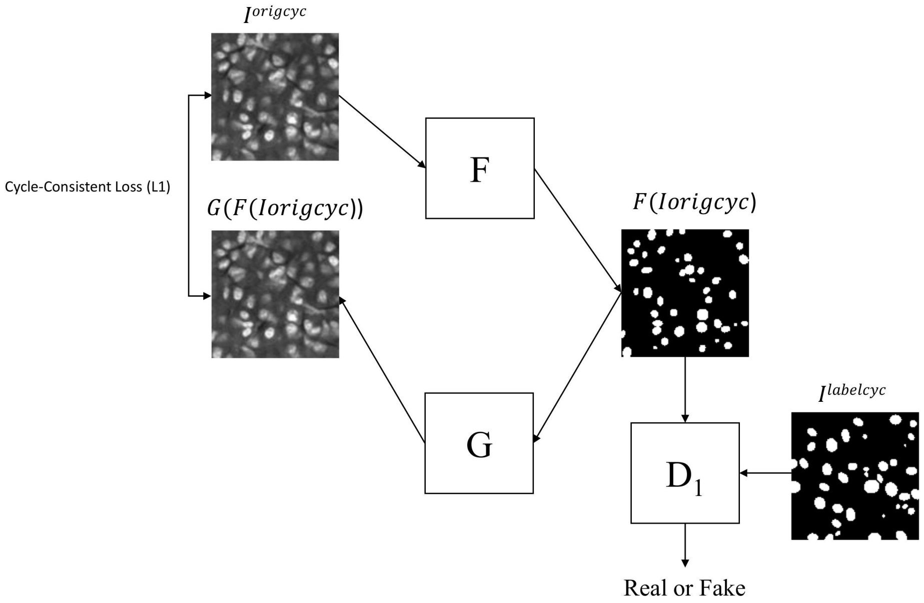


Fig. 3.5.: CycleGAN and SpCycleGAN training path two

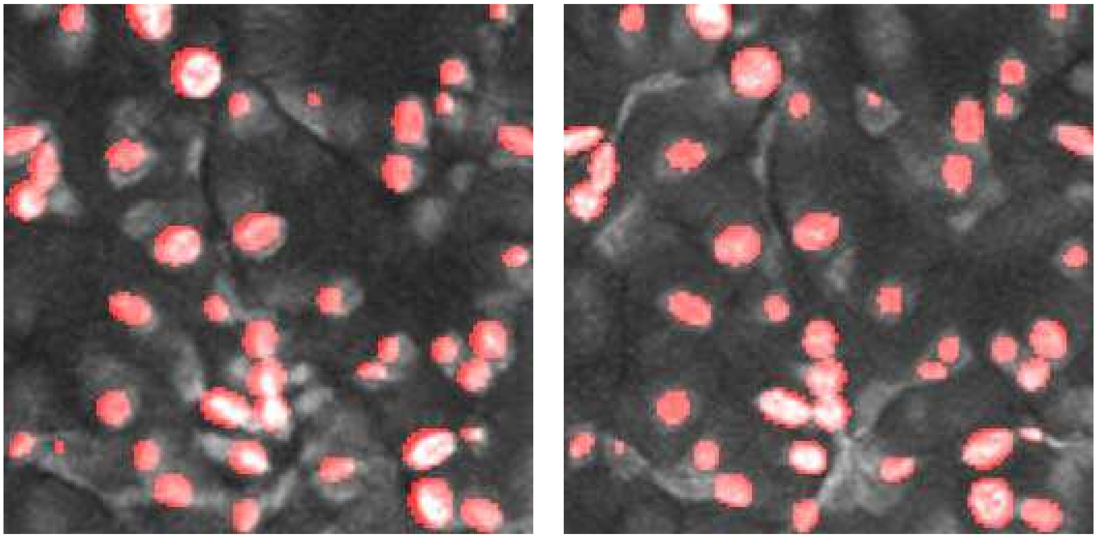
tion term in the total loss function. The network H is trained together with G. The loss function of the SpCycleGAN is defined as:

$$\begin{aligned}\mathcal{L}(G,F,H,D_1,D_2) = {} & \mathcal{L}_{\text{GAN}}(G,D_1,I^{labelcyc},I^{origcyc}) \\ & + \mathcal{L}_{\text{GAN}}(F,D_2,I^{origcyc},I^{labelcyc}) \\ & + \lambda_1\mathcal{L}_{cyc}(G,F,I^{origcyc},I^{labelcyc}) \\ & + \lambda_2\mathcal{L}_{spatial}(G,H,I^{origcyc},I^{labelcyc}) \qquad (3.4)\end{aligned}$$

where λ_1 and λ_2 are the weight coefficients for $\mathcal{L}_{cyc}$ and $\mathcal{L}_{spatial}$, respectively. Note that first three terms are the same and already defined in Equation (3.3). Here, $\mathcal{L}_{spatial}$ can be expressed as

$$\mathcal{L}_{spatial}(G,H,I^{origcyc},I^{labelcyc}) = \mathbb{E}_{I^{labelcyc}}[||H(G(I^{labelcyc})) - I^{labelcyc}||_2].$$

3.3 Experimental Results



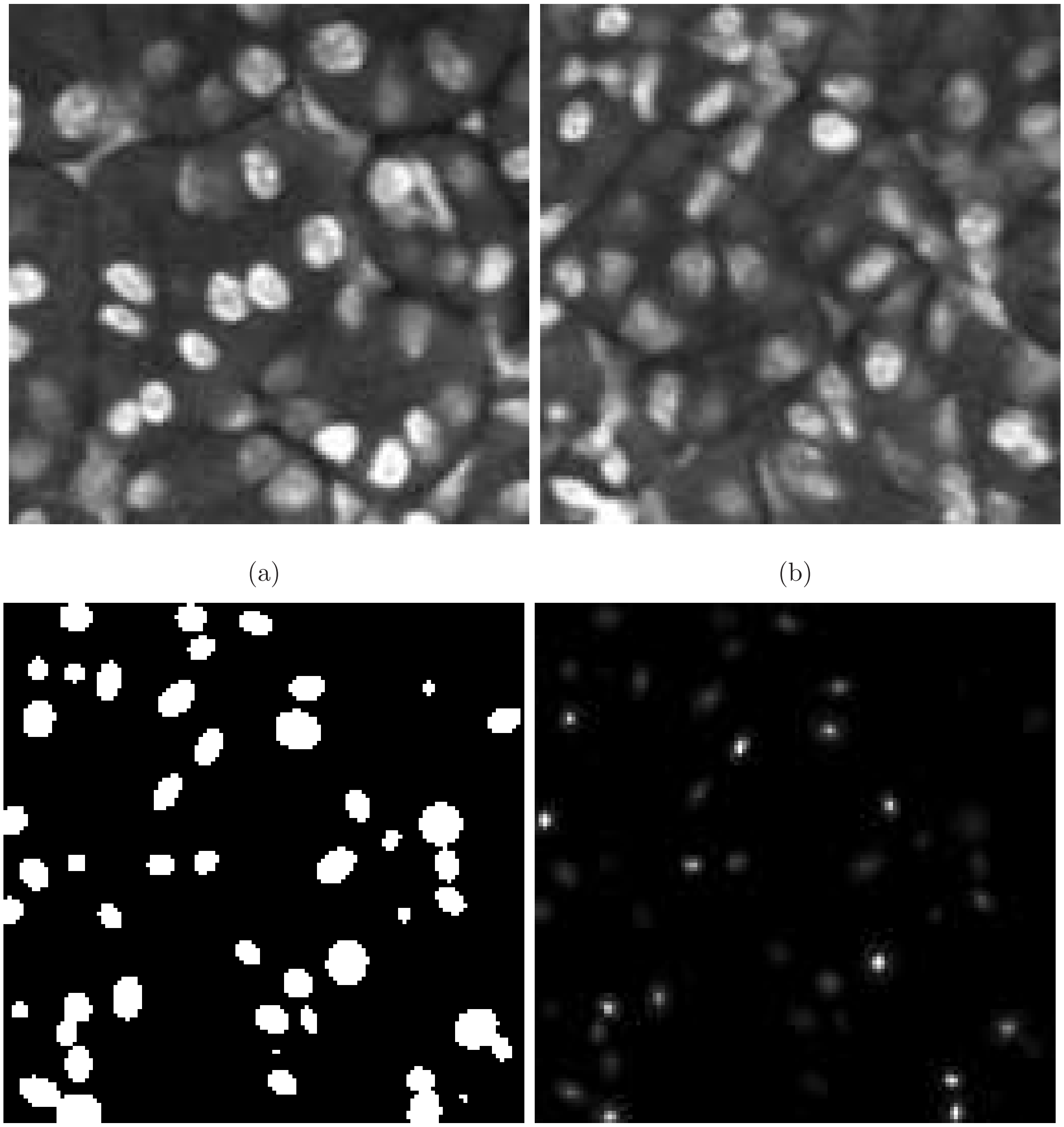
(a)

(b)

Fig. 3.6.: A comparison between two synthetic data generation methods overlaid on the corresponding synthetic binary image (a) CycleGAN, (b) SpCycleGAN

Two synthetic data generation methods between CycleGAN and SpCycleGAN from the same synthetic binary image are compared in Figure 3.6. Here, the synthetic binary image is overlaid on the synthetic microscopy image and labeled in red. It is observed that our spatial constraint loss reduces the location shift of nuclei between a synthetic microscopy image and its synthetic binary image. Our realistic synthetic microscopy volumes from SpCycleGAN can be used to train our convolutional neural networks.

As shown in figure 3.7,3.8,3.9 and 3.10, our SpCycleGAN generated synthetic images of different data sets that reflect the realistic characteristics of nuclei of original microscopy images, especially the 3D shape of nuclei. For example, the out-focused nuclei became dimmer in the I^{syn} and the corresponding nuclei in I^{label} became smaller. Most importantly, SpCycleGAN is also able to synthesize realistic non-nuclei structures which helps rejecting non-nuclei structures in training.



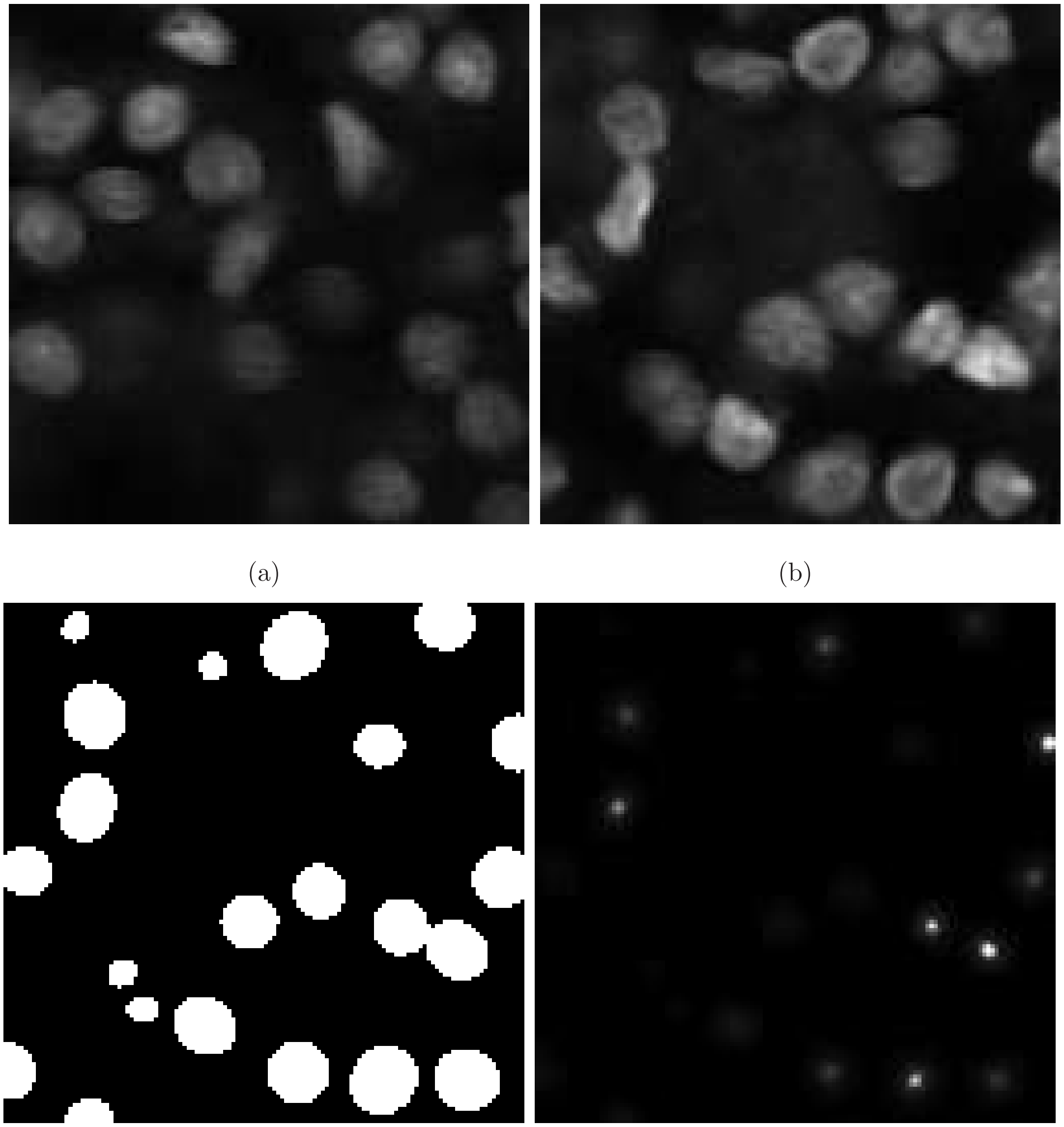
(a)

(b)

(c)

(d)

Fig. 3.7.: Synthetic data generation sample results of Data-I : (a) original microscopy image (b) images from synthetic microscopy volumes, I^{syn} (c) images from synthetic binary volumes, I^{label} (d) images from synthetic heatmap volumes, $I^{heatlabel}$



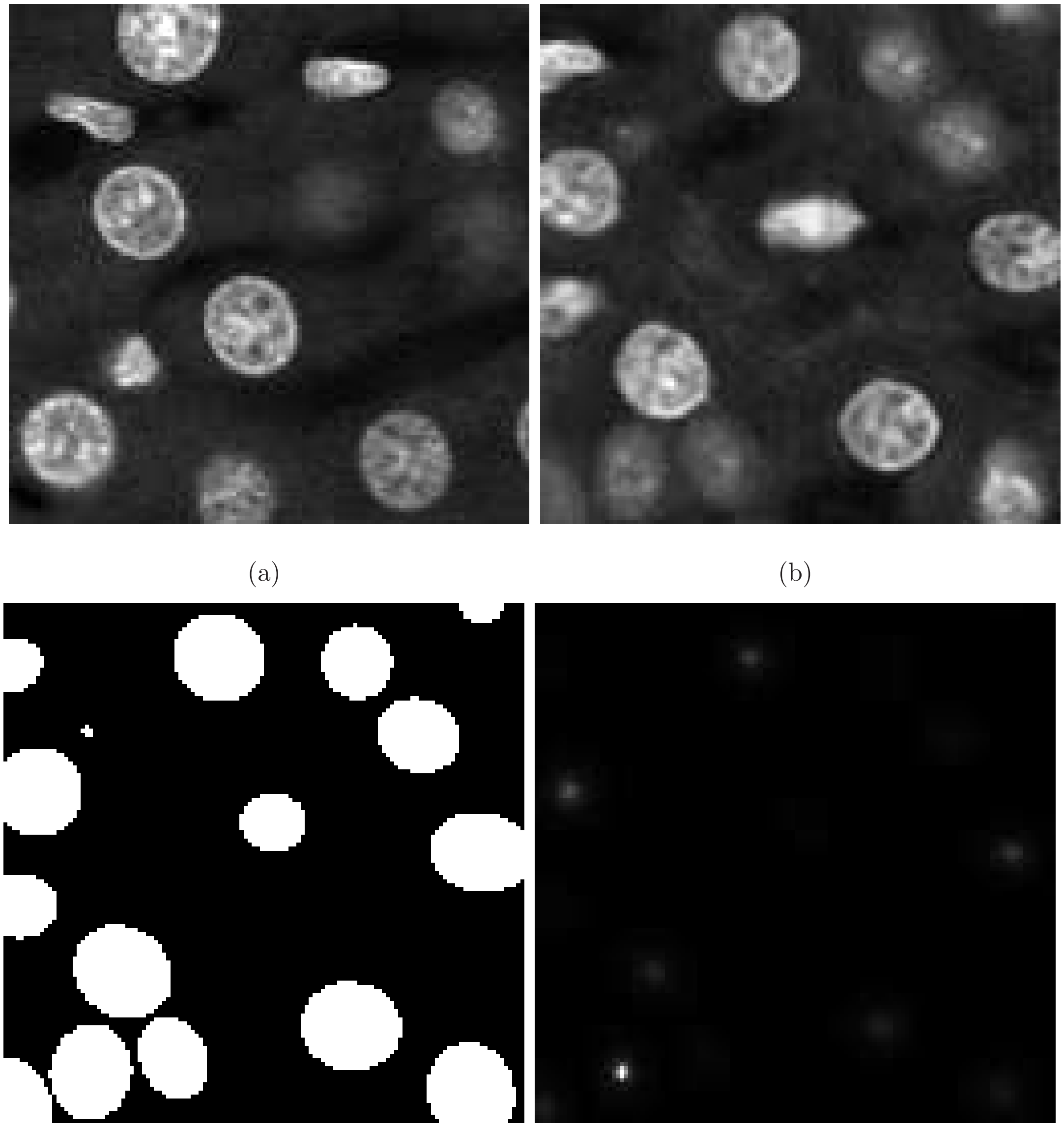
(a)

(b)

(c)

(d)

Fig. 3.8.: Synthetic data generation sample results of Data-II : (a) original microscopy image (b) images from synthetic microscopy volumes, I^{syn} (c) images from synthetic binary volumes, I^{label} (d) images from synthetic heatmap volumes, $I^{heatlabel}$



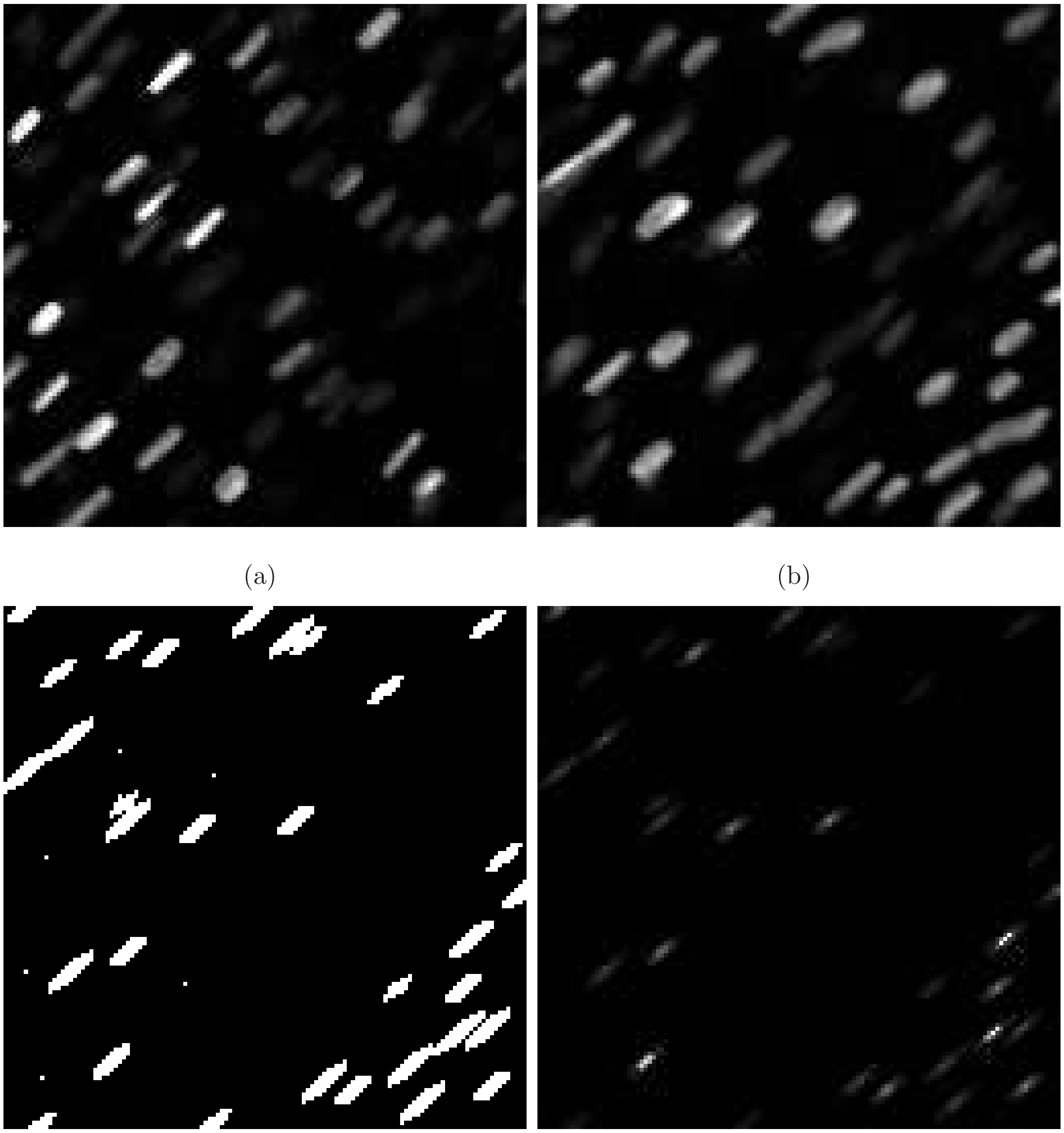
(a)

(b)

(c)

(d)

Fig. 3.9.: Synthetic data generation sample results of Data-III : (a) original microscopy image (b) images from synthetic microscopy volumes, I^{syn} (c) images from synthetic binary volumes, I^{label} (d) images from synthetic heatmap volumes, $I^{heatlabel}$



(c)

(d)

Fig. 3.10.: Synthetic data generation sample results of Data-IV : (a) original microscopy image (b) images from synthetic microscopy volumes, I^{syn} (c) images from synthetic binary volumes, I^{label} (d) images from synthetic heatmap volumes, $I^{heatlabel}$

4. CNN NUCLEI SEGMENTATION

4.1 Related Work

Fluorescence microscopy is an optical microscopy technique that has enabled the visualization of subcellular structures of in *vivo* tissue or cells [54]. Two-photon microscopy with near-infrared excitation light can now provide the ability to image deeper into tissue [10,11]. This has resulted in acquisition of large data volumes that increasingly require automated image analysis techniques for quantification studies. Image segmentation is the key step in analysis biological structures in any quantification studies. Accurate and efficient segmentation technique plays important role in biological researches. In particular, automatic image segmentation techniques are now required for the analysis of biological structures produced by fluorescence microscopy.

There have been many segmentation techniques that have been developed. Many methods are based on thresholding. Otsu's threshold [55] was proposed to automatically determine threshold that minimizes the intra-class variance and maximize inter-class variance. Instead using a global threshold proposed in Otsu's method, Niblack [56] and Sauvola [57] proposed local thresholding methods which are more useful for intensity inhomogeneity background. However, thesholding based methods have limited performance in the dataset with poor-defined structures. A active contour method was introduced in [58], which forms an initial contour and iteratively minimizes a energy function to update the contour to fit the objects. The initialization of active contour is very important in active contour since the outcome and converging time are highly depend on initial contour. The initial contour will quickly converging to fit the boundary of the region of interest if the initial contour is close to the objects. The bad initialization not only can causes long processing time but also may generates false detected masks. To overcome this problem, manual annotated initialization could be used. However, manual initialization requires a huge amount of time especially in 3D dataset. A automatic initialization technique [59] estimates the external energy field from the Poisson inverse gradient approach to generate better initial contour on the images with heavy noise. Since edge-based active contour methods [58–60] could be very sensitive to noise and depends on initial contour, region-based active contour methods are introduced to overcome the problems.

Region-based active contour approaches usually are more independent from initial contour and less sensitive to image noise. Chan-Vese 2D region-based active contour model [61] for 2D nuclei segmentation attempts to find an energy equilibrium between foreground and background region. It was extended to segment 3D cellular structures of rat kidney via region-based active surfaces in [62]. The intensity inhomogeneity and blurred edges are common in fluorescence microscopy images. [62] enhances the edge detail and solve the intensity inhomogeneity using 3D morphological filters. However, this method tends to segment multiple overlapped objects as one single object. The watershed method [63] is widely used for nuclei segmentation to solve this problem. The watershed method combines regions growing and edge detection techniques: it finds local minima (basin) and groups adjacent voxels to form a cluster and build watersheds to separate neighbor clusters. However, since the selection of local minima is highly dependent on the shape of interest and noise, watershed method tends to segment larger regions than expected. To address this over segmentation problem, marker controlled watershed was introduced in [64] that replaces local minima with predefined markers. Also, this marker controlled watershed was improved using mean shift and Kalman filter to automatically determine marker locations so that it can achieve better segmentation of time-lapse microscopy [65]. In [66], a fully automatic segmentation approach was described using multiple level set functions with a penalty term and a volume conservation constraint to separate touching cells. [66] separate touching cells by assuming the volume of cells are approximately constant. However, this assumption is not true in many cases. Besides, [66] is computationally expensive and sensitive to parameters. These were improved in [67] by using a watershed method for initialization, a non-PDE-based energy minimization for efficient computation, and the Radon transform for separating touching cells. Alternatively, [68] introduced a discrete multi-region competition method wherein the number of regions is unknown, while [69,70] implemented a method called Squassh using a energy functional derived from a generalized linear model to couple image restoration and segmentation.

A active mask method [71] segments cells uses masks instead of contours. Random initial masks were first generated and iteratively updated by region-based distributing functions and the voting-based distributing functions. Although active contour and its variations generate great results in many biomedical applications, it still has limitations. The noisy and poor image contrast are the main problem when using active contour methods.

Region competition is another segmentation technique for fluorescence image segmentation. Region competition minimizes a generalized Bayes/MDL energy function to segment objects [72]. Region competition method that can segment multiple region was developed using discrete level sets [68, 73, 74]. This method is used to segment multiple cells in fluorescence microscopy images. Squassh [69, 70] is develop to detect and quantify the subscelluar structure of fluorescence microscopy images. Squassh uses an alternating split-Bregmen algorithm to find the minimum of a energy function that derived from a generalized linear model. Squassh may has promising performance when the optimal parameters are used. However, it is hard to find a optimal parameter settings when applying with different datasets due to a large number of parameters.

Wavelets [75] also widely used for image segmentation. The wavelets for detecting edges are introduced in [75]. [76] described a method for neurons and dendrites segmentation by using a multi-scale wavelet edge detection technique. In [77], a edge detection technique using directional wavelet transform was developed.

Nonetheless, these techniques are not able to segment nuclei when other subcellular structures are present. To address this [78] developed a nuclei segmentation method that detects primitives corresponding to nuclei boundaries and delineates nuclei using region growing. This method can segments overlapped nuclei and nuclei with blurred boundary. Recently, [79] described a nuclei segmentation technique based on midpoint analysis, distance function optimization for shape fitting, and marked point processes (MPP). While producing good results, both [78] and [79] are 2D segmentation approaches that do not utilize z-directional information in a volume. Incorporating z-directional information increases the analysis complexity. [80] described a 3D cell detection and segmentation method to segment tough cells. A series of image enhancement techniques were first used on images. Touch cells are marked by seeds and then divided using a random walker algorithm. This technique utilize the physical property of cells. A assumption that the cells are generally convex is used for separating the cells. Therefore, this method may not work on objects with non-circular shapes.

Another powerful tool to segment different structures is deep learning techniques [81–83]. While deep learning methods tend to be computationally intensive, revived interest in them can be attributed to Hinton's work on *contrastive divergence* [84] as well as advancements in graphics processing units (GPUs) that have decreased the execution times.

A common deep learning approach to image classification and segmentation is based on the use of convolutional neural networks (CNN) [85–88]. The first successful application using CNN is LeNet [85] for hand-written digits recognition. [87] intro-

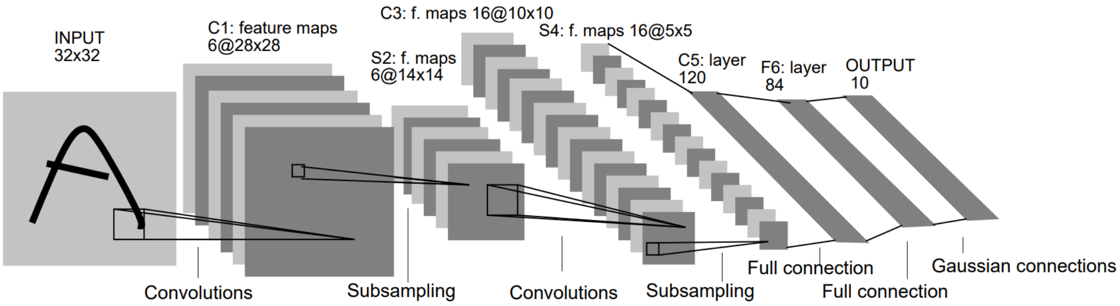


Fig. 4.1.: Architecture of LeNet

duced rectified linear unit (ReLU) to achieved best results on ImageNet classification benchmark. Nowadays, CNNs widely used in many segmentation problems. Fully convolutional networks was firstly introduced in [88] which used an encoder-decoder architecture to accomplish semantic scene segmentation. This encoder-decoder architecture was extended in SegNet [89] that utilizing VGG network architecture [90] as an encoder network and adding up corresponding decoder network with sharing pooling indices to perform better image segmentation. In [91] a CNN with max-pooling

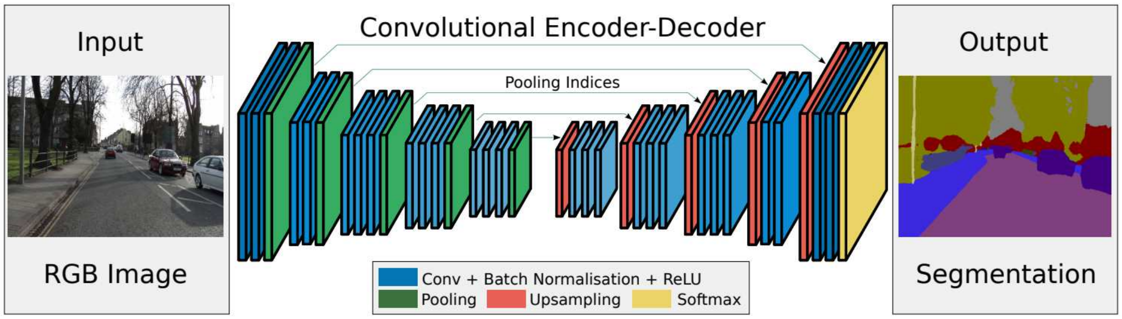


Fig. 4.2.: Architecture of Segnet

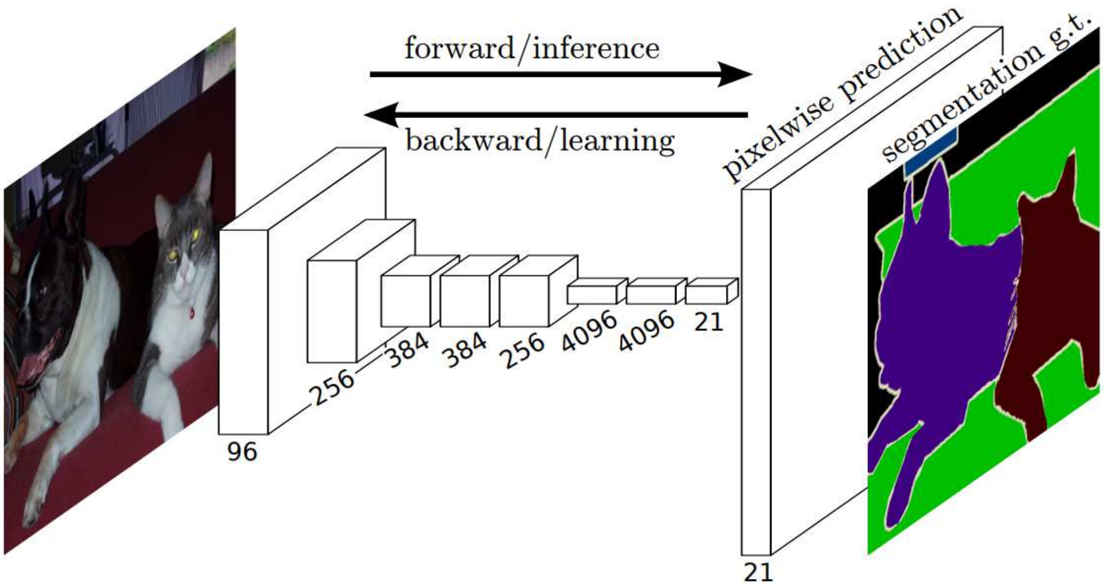


Fig. 4.3.: Architecture of Fully Convolutional Networks

layers is used to segment neuron membranes from electron microscopy images, while a CNN is used in [92] to detect Tyrosine Hydroxylase-containing cells in zebrafish images acquired using wide-field microscopy. In the case of the latter a Support Vector Machine classifier [93] was used to aid in the selection of the training data provided to the CNN. Similarly, U-Net, a fully connected convolutional network [94], has been utilized for segmentation of microscopy images. Due to the lack of training data [94] developed data augmentation methods using elastic deformations to train the CNN architecture using a small number of training images. As shown in Figure 4.4, U-Net also uses a skip connection between encoder and decoder of the architecture to preserve spatial information. In [81], a 3D U-Net is described for volumetric segmentation by expanding 2D U-Net [94]. 3D U-Net [81] uses 3D operations to fully utilize volumetric data and it requires manual annotations for training. Alternatively,

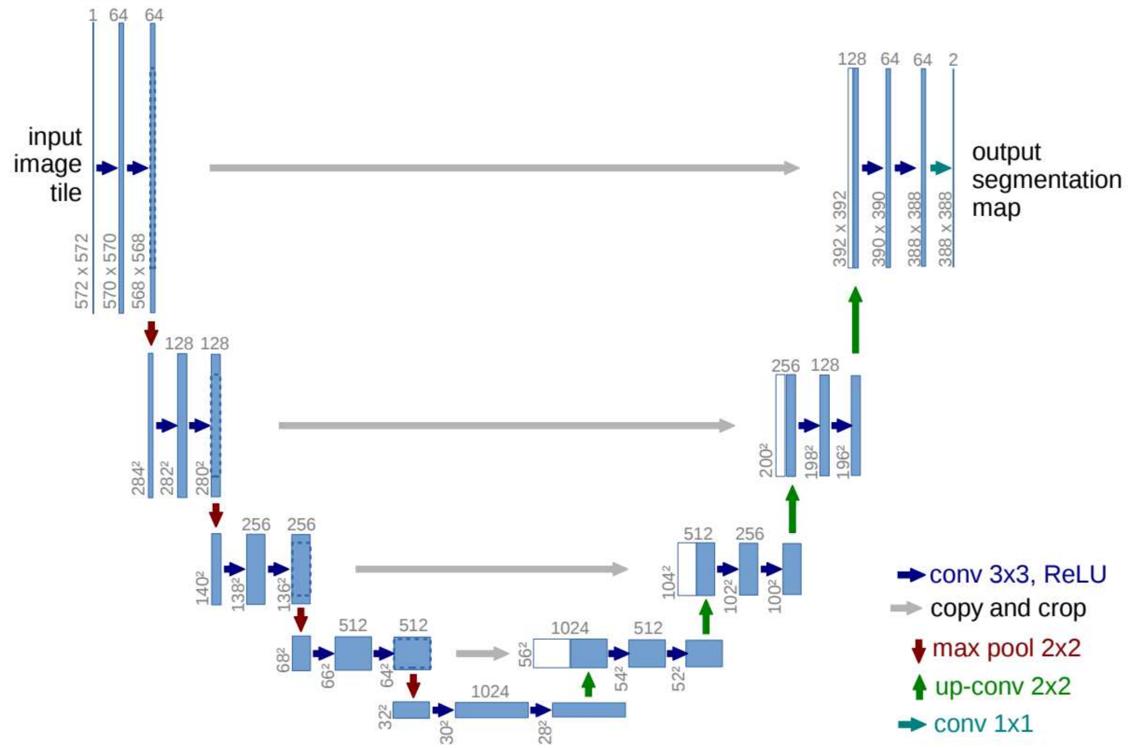


Fig. 4.4.: Architecture of U-Net

MIMO-Net which uses multiple-input and multiple-output CNN for cell segmentation in fluorescence microscopy was demonstrated in [95]. To identify overlapped nuclei, a nuclei instance segmentation method that incorprates detection and segmentation together is decribed in [96]. Another approach employs adversarial training is presented in [97]. This method [97] segments the contour of the cell to identify each individual cell.

Additionally, for nuclei segmentation of histopathology images, [98] used deep CNN-based shape initialization, whereas [99] developed a spatially constrained convolutional neural network (SC-CNN) that uses the distance from the center of a nucleus to produce a probability map. A segmentation method using a triplanar CNN [100] is described by training three two-dimensional CNNs in horizontal, frontal, and sagittal planes independently and fusing them in a final layer but it would be computationally expensive to train individual CNNs. More recently, V-Net [101] that includes the Dice loss in CNN training for volumetric medical image segmentation was described. Also, a 3D nuclei instance segmentation combining 3D nuclei detection and segmentation was demonstrated in [102].

4.2 Deep 2D+[1]

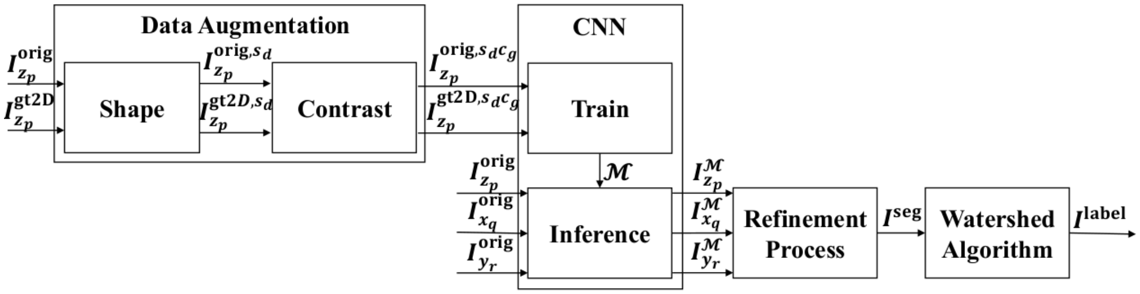


Fig. 4.5.: Block Diagram of the Proposed Method

Figure 4.5 shows a block diagram of our method. Let I be a 3D image volume of size $X \times Y \times Z$. We denote the p^{th} focal plane image, of size $X \times Y$, along the

[1] This is joint work with Dr. David J. Ho and Ms. Shuo Han.

z-direction in a stack of Z images by I_{z_p}, where $p \in \{1, \ldots, Z\}$. For example, $I_{z_{70}}^{\text{orig}}$ is the 70th original image. The volume of images can be also sliced along the x and y-directions. Let I_{x_q} be the q^{th} image, of size $Y \times Z$ in a stack of X images, along the x-direction, where $q \in \{1, \ldots, X\}$, and I_{y_r} the r^{th} image, of size $X \times Z$ in a stack of Y images, along the y-direction, where $r \in \{1, \ldots, Y\}$.

To generate training data we initially manually segment a stack of images to form "ground truth" images and then transform the segmented ground truth images using a combination of spatial and contrast transformations to form "augmented ground truth" images. Data augmentation is necessary in biomedical image analysis since ground truth images are limited due to the difficulty of manually segmenting the data. For example, in our case, only 10 ground truth images were provided. Let $I_{z_p}^{\text{gt2D}}$, denote a ground truth image version of $I_{z_p}^{\text{orig}}$, and $I_{z_p}^{\text{gt2D},s_d}$ and $I_{z_p}^{\text{gt2D},s_d c_g}$ the spatially and contrast transformed versions, respectively. This entire data set is then used to train a CNN model, $\mathcal{M}$. By testing the model, $\mathcal{M}$ on $I_{z_p}^{\text{orig}}$, $I_{x_q}^{\text{orig}}$, and $I_{y_r}^{\text{orig}}$, the segmented images, $I_{z_p}^{\mathcal{M}}$, $I_{x_q}^{\mathcal{M}}$, and $I_{y_r}^{\mathcal{M}}$, are generated, respectively.

This is followed by a refinement process, whose outcome, I^{seg}, consists of 3D nuclei. Finally, a watershed technique [63, 103] is used to label individual nuclei so they can be counted. The result of the watershed technique is denoted by I^{label}.

4.2.1 Proposed Data Augmentation Approach

Shape Augmentation

Since fluorescence microscopy images contain nuclei of various shapes, using elastic deformations for data augmentation yields better performance than random rotations, shifts, shear, and flips. An elastic deformation produces different shapes and orientations of nuclei by warping images locally. In order to change the shapes and orientations of nuclei in $I_{z_p}^{\text{orig}}$ and $I_{z_p}^{\text{gt2D}}$, a grid of control points with 64 pixel spacing in the x and y-directions is created for each image. Control points are then randomly displaced in both the x and y-directions to within ±15 pixels. A two-dimensional

B-spline is then fit to the grid of displaced control points, and all pixels in $I_{z_p}^{\text{orig},s_d}$, and its ground truth version, $I_{z_p}^{\text{gt2D},s_d}$, are displaced accordingly. We use bicubic interpolation to warp each pixel to its new coordinates. This process is repeated to produce 100 randomly generated images,indexed by d, from one image where $d \in \{1, \ldots, 100\}$.

Contrast Augmentation

Fluorescence microscopy images generally have inhomogeneous intensities and boundary regions tend to be darker than the central regions. To segment nuclei in both darker regions and brighter regions, it is necessary to train the model with various contrast between nuclei and other structures. Hence, for each $I_{z_p}^{\text{orig},s_d}$, we use a contrast transformation:

$$v = 255\left(\frac{u}{255}\right)^{\frac{1}{\gamma}} \tag{4.1}$$

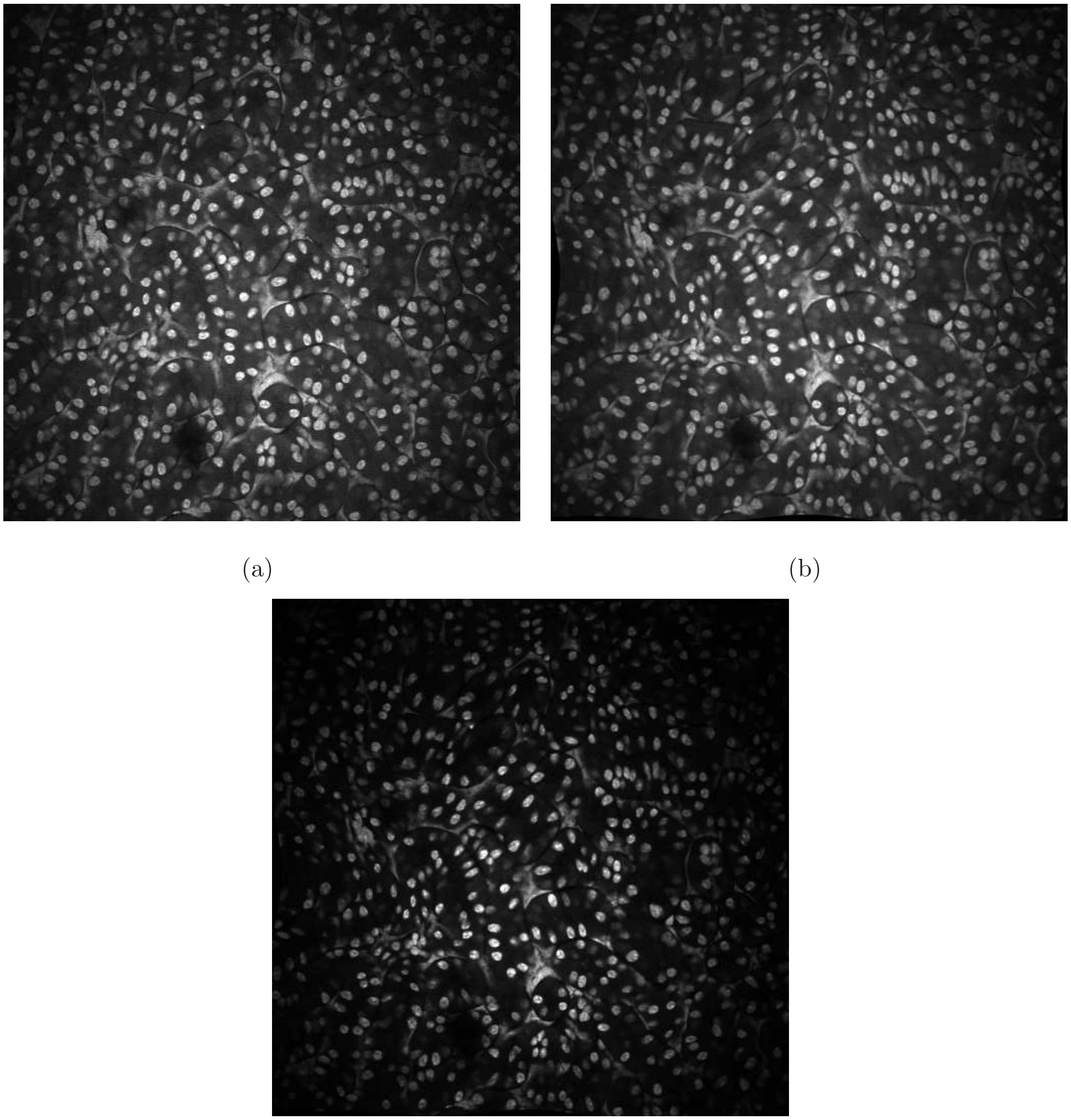
,where u denotes a pixel value in $I_{z_p}^{\text{orig},s_d}$ and v the corresponding pixel value in the contrast transformed $I_{z_p}^{\text{orig},s_d c_g}$. Here g is a parameter such that

$$\gamma = \frac{\log(\frac{1}{2})}{\log(\frac{g}{255})} \tag{4.2}$$

. Note that values of $g \leq 127$ will generate a darker image and $g \geq 128$ will produce a brighter image than the original image. In our implementation for each $I_{z_p}^{\text{orig},s_d}$, 9 more images with different contrast parameters g, $I_{z_p}^{\text{orig},s_d c_g}$, and corresponding ground truth images, $I_{z_p}^{\text{gt2D},s_d c_g}$, are generated where $g \in \{80, 90, \ldots, 160\}$. Figure 4.6 depicts images generated using the data augmentation approach.

4.2.2 Convolutional Neural Network (CNN)

The architecture of our convolutional neural network, shown in Figure 4.7, is based on the architecture of SegNet which was originally designed for road scene segmentation [89]. To adapt the architecture to nuclei segmentation the deepest layer in SegNet is removed, and four encoder-decoder layers employed. Each encoder layer consists of a combination of multiple convolutional layers and a pooling layer,



(a)

(b)

(c)

Fig. 4.6.: Proposed data augmentation approach generates multiple training images (a) $I_{z_{70}}^{\mathrm{orig}}$, (b) $I_{z_{70}}^{\mathrm{orig},s_1}$, (c) $I_{z_{70}}^{\mathrm{orig},s_1 c_{80}}$

batch normalization [104], and a rectified-linear unit (ReLU) activation function. Additionally, each convolutional layer performs a convolutional operation with a 3×3 kernel with 1 pixel padding being used to maintain the original size of the images. In

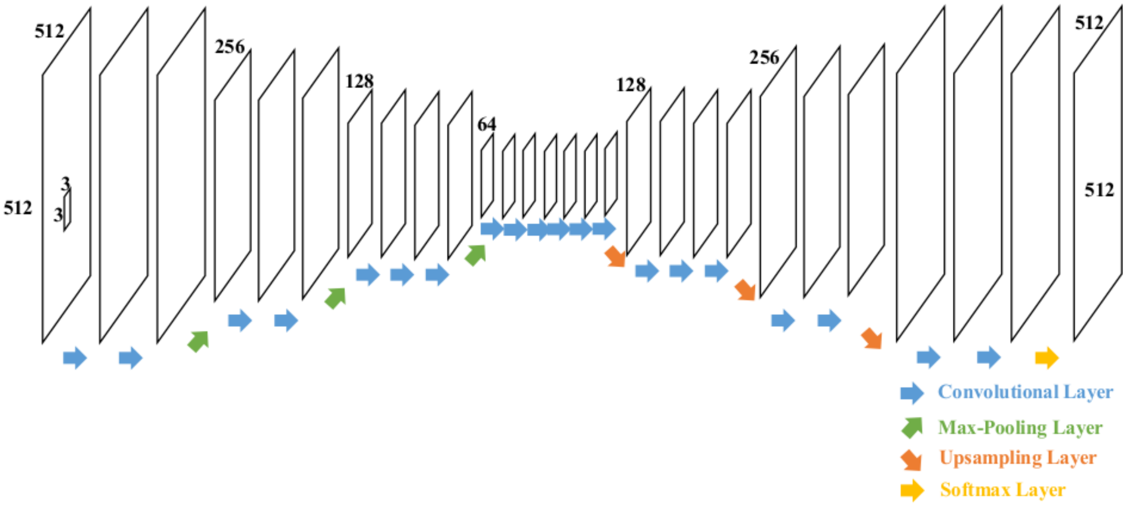


Fig. 4.7.: Architecture of our convolutional neural network

the pooling layer, images are down sampled by a 2 × 2 max pooling operation with stride of 2.

Conversely, a decoder consists of an up sampling layer and multiple convolutional layers, where the convolutional layers in the decoder have the same structure as the encoder convolutional layers. This is followed by the final layer of the architecture, where a softmax classifier is used to estimate the probability that a pixel belongs to a nucleus or not.

To train our model, $\mathcal{M}$, we used the Caffe implementation [105] of stochastic gradient descent (SGD) with a fixed learning rate of 0.01 and a momentum of 0.9. For each iteration, a randomly-selected training image is used to train $\mathcal{M}$. In the softmax classifier, class weights of nuclei and background (non-nuclei) are pre-set according to the number of pixels in each class divided by the total number of pixels in images.

4.2.3 Refinement Process

Based on the current architecture, $\mathcal{M}$ can only segment 2D images. To extend it to 3D we use $\mathcal{M}$, not only with $I_{z_p}^{\text{orig}}$ but also with $I_{x_q}^{\text{orig}}$ and $I_{y_r}^{\text{orig}}$, denoting the segmented results as $I_{z_p}^{\mathcal{M}}$, $I_{x_q}^{\mathcal{M}}$, and $I_{y_r}^{\mathcal{M}}$, respectively. For each voxel in a volume, (q, r, p), the final label, $I^{\text{seg}}(q, r, p)$, is then selected by performing a majority vote on the voxel from the stack of $I_{z_p}^{\mathcal{M}}(q, r)$, $I_{x_q}^{\mathcal{M}}(r, p)$, and $I_{y_r}^{\mathcal{M}}(q, p)$. For example, for voxel $(50, 60, 70)$, the final label, $I^{\text{seg}}(50, 60, 70)$, will considered to be a nuclei if two or more objects identified in $I_{z_{70}}^{\mathcal{M}}(50, 60)$, $I_{x_{50}}^{\mathcal{M}}(60, 70)$, and $I_{y_{60}}^{\mathcal{M}}(50, 70)$ are labeled as nuclei. Otherwise, $I^{\text{seg}}(50, 60, 70)$ will be labeled as background.

4.2.4 Watershed Based Nuclei Separation and Counting

In biological samples, it is common for two or more nuclei to be located closely, or even overlap. Moreover, photons emitted by nuclei in one focal plane can still be detected in the current focal plane resulting in nuclei that appear to be elongated and interconnected. We use a 3D watershed technique [63, 103] as a post processing technique to demarcate individual nucleus since watershed is effective in separating multiple overlapping objects. For example, [65] uses marker-controlled watershed to segment touching nuclei. Watershed generates a distinct label for each nucleus by finding local minima in the topographical distance transform [103]. If it is used on an original volume, local minima will be assigned not only to nuclei but also to other structures. Since I^{seg} contains segmented nuclei, the watershed demarcates overlapping nuclei and labels adjacent nuclei. The total number of nuclei in the volume can thus be estimated based on the number of labels. This information is particularly useful for analyzing properties such as cell livability for biological studies.

4.2.5 Experimental Results

Our method was tested on four different rat kidney data sets. All data sets consisted of $X = 512 \times Y = 512$ sized grayscale images. Data-I consists of $Z = 512$ images, Data-II of $Z = 36$ images, Data-III of $Z = 23$ images, and Data-IV of $Z = 45$. To train the model, $\mathcal{M}$, we only used 10 ground truth images in z-direction, $I_{z_p}^{\text{gt2D}}$, from Data-I.

We cropped the entire volume of Data-I into three sub-volumes: Volume-I, Volume-II, and Volume-III each of size $32 \times 32 \times 32$ to visualize the results. Volume-I, Volume-II, and Volume-III have the same x and y coordinates, $241 \leq q \leq 272$ and $241 \leq r \leq 272$, but different depths, where $31 \leq p \leq 62$ for Volume-I, $131 \leq p \leq 162$ for Volume-II, and $231 \leq p \leq 262$ for Volume-III.

Our method was evaluated using 3D ground truth volumes, I^{gt3D}, with size of $32 \times 32 \times 32$ and its performance is compared with other known methods, as shown in Figure 4.8, using Voxx [106]. In order to generate I^{gt3D}, we first generated three sets of cropped 2D ground truth images with size of 32×32 on $I_{z_p}^{\text{orig}}$, $I_{x_q}^{\text{orig}}$, and $I_{y_r}^{\text{orig}}$, respectively. By using the refinement process on the cropped 2D ground truth images, 3D ground truth volumes are generated. During the 3D ground truth generation, we are visually helped by Voxx [106].

As shown in Figure 4.8 (c) and (d), 3D active surfaces from [62] and 3D Squassh [69, 70] have no ability to distinguish nuclei from other subcellular structures. Also, from Figure 4.8 (e) and (f), our refinement process segments nuclei into more ellipsoidal shape since it combines horizontal, frontal, and sagittal information.

In addition to the visual evaluation, the accuracy, Type-I error, and Type-II error metrics were used to evaluate segmentation performance. Here, accuracy $= \frac{n_{\text{TP}}+n_{\text{TN}}}{n_{\text{total}}}$, Type-I error $= \frac{n_{\text{FP}}}{n_{\text{total}}}$, Type-II error $= \frac{n_{\text{FN}}}{n_{\text{total}}}$, where n_{TP}, n_{TN}, n_{FP}, n_{FN}, n_{total} denote the number of true-positives (voxels belonging to nuclei that are correctly segmented as nuclei), true-negatives (voxels belonging to background that are correctly segmented as background), false-positives (voxels belonging to background that are incorrectly segmented as nuclei), false-negatives (voxels belonging to nuclei that are incorrectly segmented as background), and the total number of voxels in a volume, respectively. As shown in Table 4.1, our method outperformed the other methods.

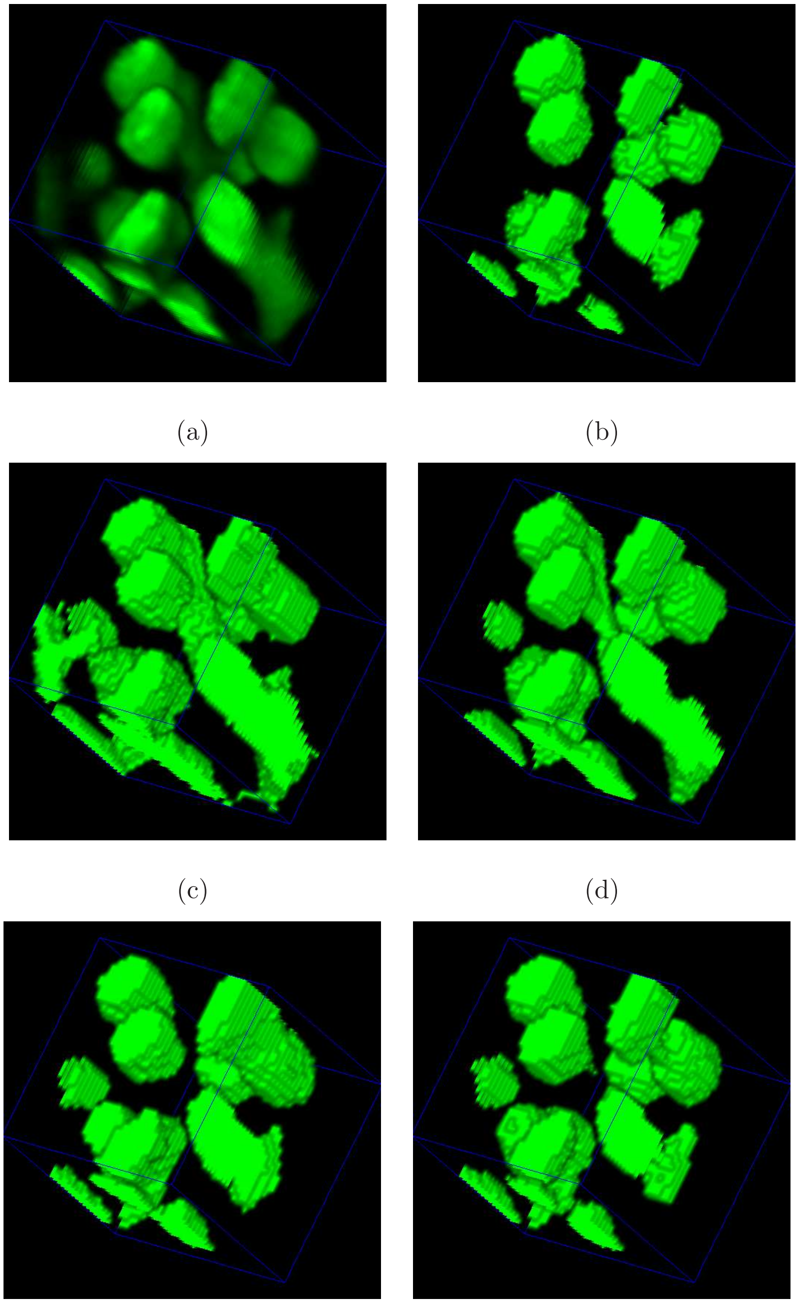
Table 4.1.: Accuracy, Type-I and Type-II errors for other methods and our method on the Data-I

		Method [62]	Method [69, 70]	Our Method
Vol.-I	Acc.	84.62%	90.14%	**94.25%**
	Type-I	14.80%	9.07%	5.18%
	Type-II	0.25%	0.79%	0.57%
Vol.-II	Acc.	79.67%	88.26%	**95.24%**
	Type-I	20.16%	11.67%	4.18%
	Type-II	0.16%	0.07%	0.58%
Vol.-III	Acc.	76.72%	87.29%	**93.21%**
	Type-I	23.24%	12.61%	6.61%
	Type-II	0.05%	0.10%	0.18%

The experimental results indicated that our method produces accurate boundaries for nuclei and provides visual assistance for researchers to identify individual nucleus. Figure 4.9 compares the original image, the CNN based result, and the result of watershed segmentation on the result of the CNN.

In addition, we compare the estimated nuclei count with count of ground truth data in 25 tissue volumes of $32 \times 32 \times 32$ in size. The mean of the ground truth nuclei count in the 25 volumes is 15.32 ± 3.76 whereas the mean of the estimated nuclei count using the proposed method is 15.84 ± 3.91. A two-tailed t-test indicates the difference between the estimate and ground truth is not statistically significant (having a p-value of p=0.634).

The trained model based on Data-I can also successfully segment nuclei from different rat kidney data (Data-II, Data-III, Data-IV) as shown in Figure 4.10. Note



(a)

(b)

(c)

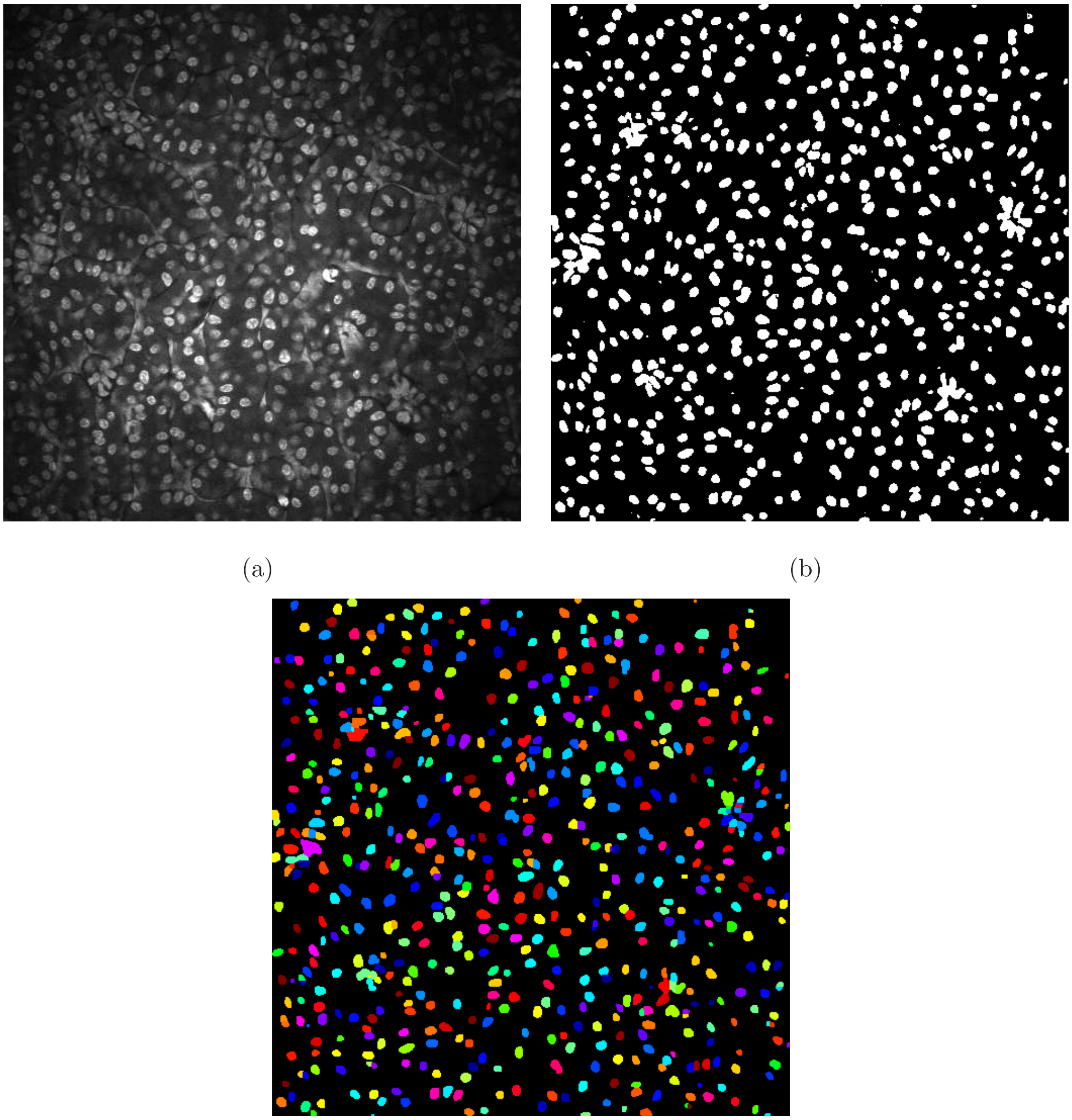
(d)

(e)

(f)

Fig. 4.8.: 3D visualization of Volume-I of Data-I using Voxx [106] (a) original volume (b) 3D ground truth volume, (c) 3D active surfaces from [62], (d) 3D Squassh from [69, 70], (e) segmentation result before refinement, (f) segmentation result from after refinement.

that our model, $\mathcal{M}$, is trained from ground truth images from only Data-I. Although the size of nuclei presented in Data-II, Data-III, and Data-IV is different from the size of nuclei in Data-I, our technique can still segment and count nuclei in data-II, Data-III, and Data-IV.

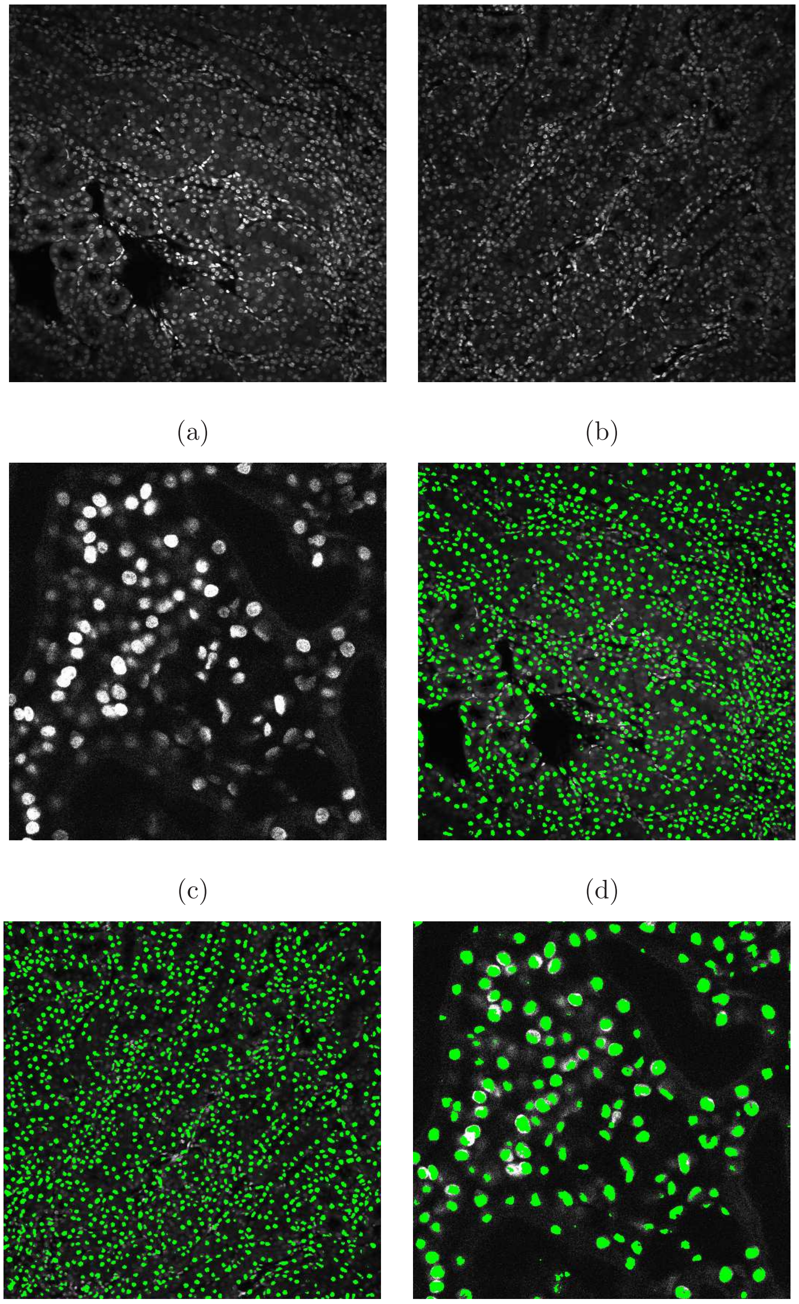


(a)

(b)

(c)

Fig. 4.9.: Nuclei count using watershed (a) original image, $I_{z_{175}}^{\text{orig}}$, (b) segmentation result from our method, $I_{z_{175}}^{\text{seg}}$, (c) watershed result, $I_{z_{175}}^{\text{label}}$



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 4.10.: Nuclei segmentation on different rat kidney data (a) $I^{\text{orig}}_{z_{16}}$ of Data-II, (b) $I^{\text{orig}}_{z_{13}}$ of Data-III, (c) $I^{\text{orig}}_{z_{23}}$ of Data-IV, (d) $I^{\text{seg}}_{z_{16}}$ of Data-II, (e) $I^{\text{seg}}_{z_{13}}$ of Data-III, (f) $I^{\text{seg}}_{z_{23}}$ of Data-IV

4.3 Deep 3D+

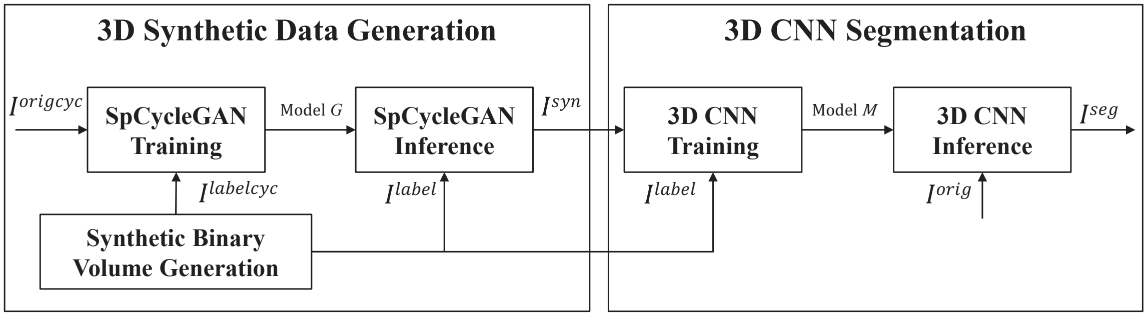


Fig. 4.11.: Block Diagram of Deep 3D+

Figure 4.11 shows a block diagram of our method. We denote I as a 3D image volume of size $X \times Y \times Z$. Note that I_{z_p} is a p^{th} focal plane image, of size $X \times Y$, along the z-direction in a volume, where $p \in \{1, \ldots, Z\}$. Note also that I^{orig} and I^{seg} is the original fluorescence microscopy volume and segmented volume, respectively. In addition, let $I_{(q_i:q_f, r_i:r_f, p_i:p_f)}$ be a subvolume of I, whose x-coordinate is $q_i \leq x \leq q_f$, y-coordinate is $r_i \leq y \leq r_f$, z-coordinate is $p_i \leq z \leq p_f$, where $q_i, q_f \in \{1, \ldots, X\}$, $r_i, r_f \in \{1, \ldots, Y\}$, $p_i, p_f \in \{1, \ldots, Z\}$, $q_i \leq q_f$, $r_i \leq r_f$, and $p_i \leq p_f$. For example, $I^{seg}_{(241:272,241:272,131:162)}$ is a subvolume of a segmented volume, I^{seg}, where the subvolume is cropped between 241st slice and 272nd slice in x-direction, between 241st slice and 272nd slice in y-direction, and between 131st slice and 162nd slice in z-direction.

As shown in Figure 4.11, our proposed method consists of two steps: 3D synthetic data generation and 3D CNN segmentation. We first generate synthetic binary volumes, $I^{labelcyc}$, and then use them with a subvolume of the original image volumes, $I^{origcyc}$, to train a spatially constrained CycleGAN (SpCycleGAN) and obtain a generative model denoted as model G. This model G is used with another set of synthetic binary volume, I^{label}, to generate corresponding synthetic 3D volumes, I^{syn}. For 3D CNN segmentation, we can utilize these paired I^{syn} and I^{label} to train a 3D CNN and obtain model M. Finally, the 3D CNN model M is used to segment nuclei in I^{orig} to produce I^{seg}.

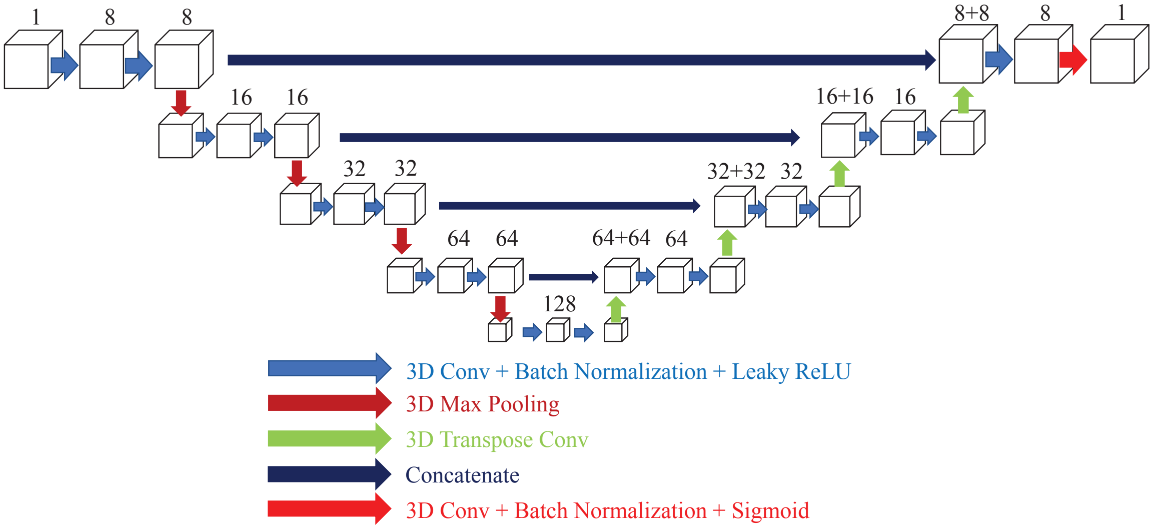


Fig. 4.12.: Architecture of our modified 3D U-Net

4.3.1 3D U-Net

Figure 4.12 shows the architecture of our modified 3D U-Net. The filter size of each 3D convolution is $3 \times 3 \times 3$. To maintain the same size of volume during 3D convolution, a voxel padding of $1 \times 1 \times 1$ is used in each convolution. A 3D batch normalization [104] and a leaky rectified-linear unit activation function are employed after each 3D convolution. In the downsampling path, a 3D max pooling uses $2 \times 2 \times 2$ with stride of 2 is used. In the upsampling path, feature information is retrieved using 3D transpose convolutions. Our modified 3D U-Net is one layer deeper than conventional U-Net as can be seen in Figure 4.12. Our training loss function can be expressed as a linear combination of the Dice loss ($\mathcal{L}_{Dice}$) and the binary cross-entropy loss ($\mathcal{L}_{BCE}$) such that

$$\mathcal{L}_{seg}(T, S) = \mu_1 \mathcal{L}_{Dice}(T, S) + \mu_2 \mathcal{L}_{BCE}(T, S) \tag{4.3}$$

where

$$\begin{aligned}\mathcal{L}_{Dice}(T,S) &= \frac{2(\sum_{i=1}^{N} t_i s_i)}{\sum_{i=1}^{N} t_i^2 + \sum_{i=1}^{N} s_i^2} \\ \mathcal{L}_{BCE}(T,S) &= -\frac{1}{N}\sum_{i=1}^{N} t_i \log(s_i) + (1-t_i)\log(1-s_i),\end{aligned}$$

respectively [101]. Note that T is the set of the targeted groundtruth values and $t_i \in T$ is a targeted groundtruth value at i^{th} voxel location. Similarly, S is a probability map of binary volumetric segmentation and $s_i \in S$ is a probability map at i^{th} voxel location. Lastly, N is the number of entire voxels and μ_1, μ_2 serve as the weight coefficient between to loss terms in Equation (4.3). The network takes a grayscale input volume with size of $64 \times 64 \times 64$ and produces an voxelwise classified 3D volume with the same size of the input volume. To train our model M, V pairs of synthetic microscopy volumes, I^{syn}, and synthetic binary volumes, I^{label}, are used.

4.3.2 Inference

For the inference step we first zero-padded I^{orig} by 16 voxels on the boundaries. A 3D window with size of $64 \times 64 \times 64$ is used to segment nuclei. Since the zero padded I^{orig} is bigger than the 3D window, the 3D windows is slided to x, y, and z-directions by 32 voxels on zero-padded I^{orig} [43]. Nuclei partially observed on boundaries of the 3D window may not be segmented correctly. Hence, only the central subvolume of the output of the 3D window with size of $32 \times 32 \times 32$ is used to generate the corresponding subvolume of I^{seg} with size of $32 \times 32 \times 32$. This process is done until the 3D window maps an entire volume.

4.3.3 Experimental Results

We tested our proposed method on two different rat kidney data sets. These data sets contain grayscale images of size $X = 512 \times Y = 512$. Data-I consists of $Z = 512$ images, Data-II consist of $Z = 64$.

Our SpCycleGAN is implemented in Pytorch using the Adam optimizer [107] with default parameters given by CycleGAN [51]. In addition, we used $\lambda_1 = \lambda_2 = 10$ in the SpCycleGAN loss function shown in Equation (3.4). We trained the CycleGAN and SpCycleGAN to generate synthetic volumes for Data-I and Data-II, respectively. A $128 \times 128 \times 128$ synthetic binary volume for Data-I denoted as $I^{labelcyc_{Data-I}}$ and a $128 \times 128 \times 300$ subvolume of original microscopy volume of Data-I denoted as $I^{origcyc_{Data-I}}$ were used to train model G^{Data-I}. Similarly, a $128 \times 128 \times 128$ synthetic binary volume for Data-II denoted as $I^{labelcyc_{Data-II}}$ and a $128 \times 128 \times 32$ subvolume of original microscopy volume of Data-II denoted as $I^{origcyc_{Data-II}}$ were used to train model $G^{Data-II}$.

We generated 200 sets of $128 \times 128 \times 128$ synthetic binary volumes, $I^{label_{Data-I}}$ and $I^{label_{Data-II}}$ where $I^{label_{Data-I}}$ and $I^{label_{Data-II}}$ are generated according to different size of nuclei in Data-I and Data-II, respectively. By using the model G^{Data-I} on $I^{label_{Data-I}}$, 200 pairs of synthetic binary volumes, $I^{label_{Data-I}}$, and corresponding synthetic microscopy volumes, $I^{syn_{Data-I}}$, of size of $128 \times 128 \times 128$ were obtained. Similarly, by using model $G^{Data-II}$ on $I^{label_{Data-II}}$, 200 pairs of $I^{label_{Data-II}}$ and corresponding $I^{syn_{Data-II}}$, of size of $128 \times 128 \times 128$ were obtained. Since our modified 3D U-Net architecture takes volumes of size of $64 \times 64 \times 64$, we divided $I^{label_{Data-I}}$, $I^{syn_{Data-I}}$, $I^{label_{Data-II}}$, and $I^{syn_{Data-II}}$ into adjacent non overlapping $64 \times 64 \times 64$. Thus, we have 1600 pairs of synthetic binary volumes and corresponded synthetic microscopy volumes per each data to train our modified 3D U-Net. Note that these 1600 synthetic binary volumes per each data are used as groundtruth volumes to be paired with corresponding synthetic microscopy volumes. Model M^{Data-I} and $M^{Data-II}$ are then generated.

Our modified 3D U-Net is implemented in Pytorch using the Adam optimizer [107] with learning rate 0.001. For the evaluation purpose, we use different settings of using 3D synthetic data generation methods (CycleGAN or SpCycleGAN), different number of pairs of synthetic training volume V ($V = 80$ or $V = 1600$) among 1600 pairs of synthetic binary volume corresponding synthetic microscopy volume. Also, we use different loss functions with different settings of the μ_1 and μ_2. Moreover, we also compared our modified 3D U-Net with 3D encoder-decoder architecture [43]. Lastly, small objects which are less than 100 voxels were removed using 3D connected components.

Figure 4.13 shows the synthetic images generated by our proposed method. The left column indicates original images whereas middle column shows synthetic images artificially generated from corresponding synthetic binary images provided in right column. As can be seen from Figure 4.13, the synthetic images reflect characteristics of the original microscopy images such as background noise, nuclei shape, orientation and intensity.

Our proposed method was compared to other 3D segmentation methods including 3D active surface [62], 3D active surface with inhomogeneity correction [108], 3D Squassh [69, 70], 3D encoder-decoder architecture [43], 3D encoder-decoder architecture with CycleGAN. Three original 3D subvolumes of Data-I were selected to evaluate the performance of our proposed method. We denote the original volume as subvolume 1 ($I^{orig}_{(241:272,241:272,31:62)}$), subvolume 2 ($I^{orig}_{(241:272,241:272,131:162)}$), and subvolume 3 ($I^{orig}_{(241:272,241:272,231:262)}$), respectively. Corresponding groundtruth of each subvolume was hand segmented. Voxx [106] was used to visualize the segmentation results in 3D and compared to the manually annotated volumes. In Figure 4.14, 3D visualizations of the hand segmented subvolume 1 and the corresponding segmentation results for various methods were presented. As seen from the 3D visualization in Figure 4.14, our proposed method shows the best performance among presented methods visually compared to hand segmented groundtruth volume. In general, our proposed method captures only nuclei structure whereas other presented methods falsely detect non-nuclei structures as nuclei. Note that segmentation results in Figure 4.14(g) yields smaller segmentation mask and suffered from location shift. Our proposed method shown in Figure 4.14(h) outperforms Figure 4.14(g) since our proposed method uses spatially constrained CycleGAN and takes consideration of the Dice loss and the binary cross-entropy loss.

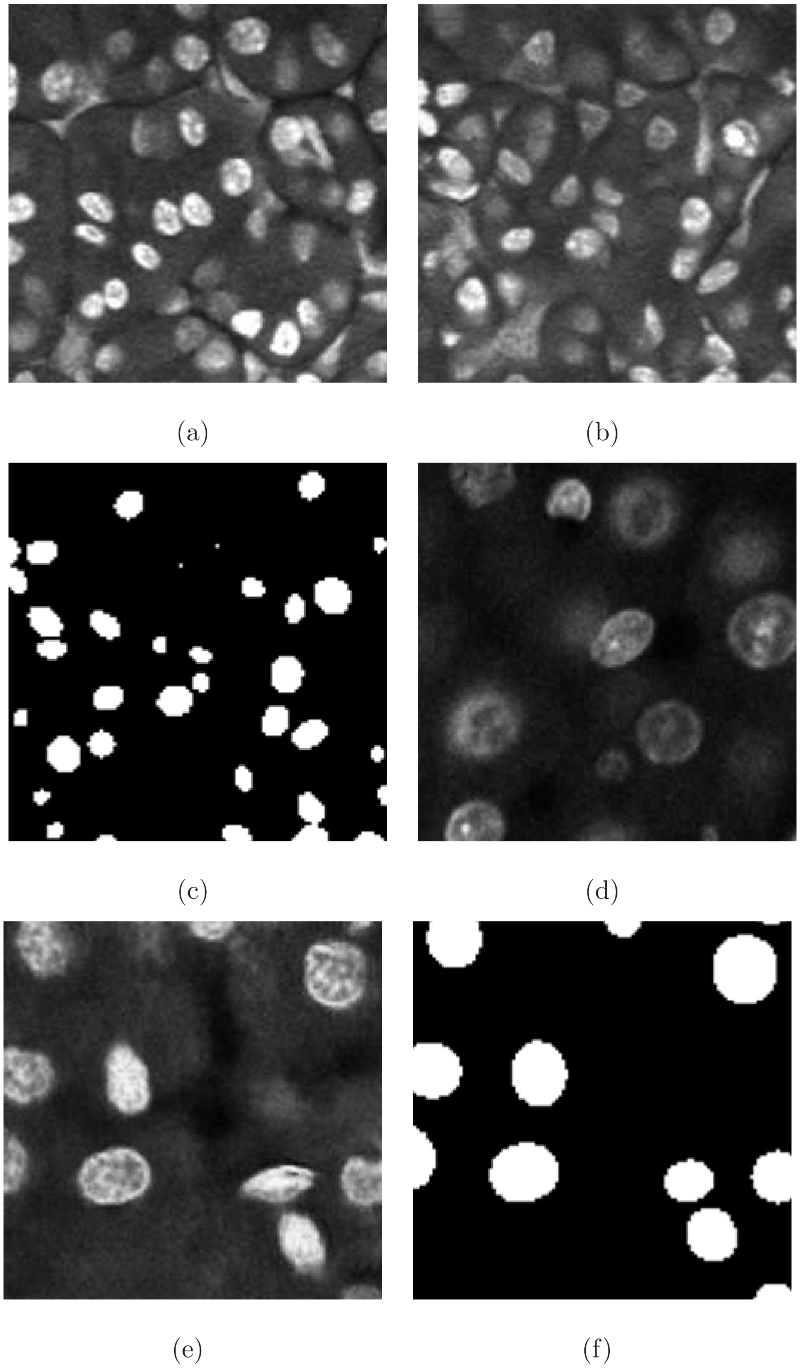


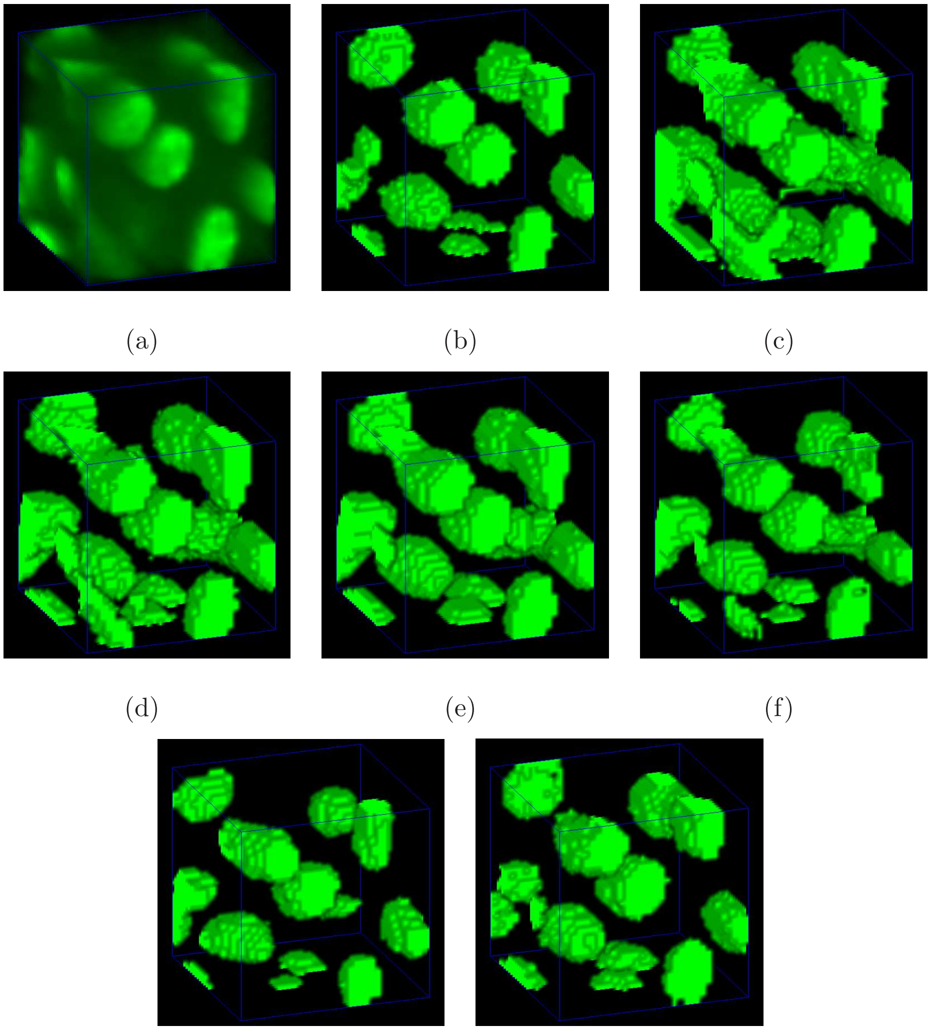
Fig. 4.13.: Slices of the original volume, the synthetic microscopy volume, and the corresponding synthetic binary volume for Data-I and Data-II (a) original image of Data-I, (b) synthetic microscopy image of Data-I, (c) synthetic binary image of Data-I, (d) original image of Data-II, (e) synthetic microscopy image of Data-II, (f) synthetic binary image of Data-II

Table 4.2.: Accuracy, Type-I and Type-II errors for known methods and our method on subvolume 1 of Data-I

	Subvolume 1		
Method	Accuracy	Type-I	Type-II
Method [62]	84.09%	15.68%	0.23%
Method [108]	87.36%	12.44%	0.20%
Method [69, 70]	90.14%	9.07%	0.79%
Method [43]	92.20%	5.38%	2.42%
3D Encoder-Decoder + CycleGAN + BCE ($\mu_1 = 0$, $\mu_2 = 1$,$V = 80$)	93.05%	3.09%	3.87%
3D Encoder-Decoder + SpCycleGAN + BCE ($\mu_1 = 0$, $\mu_2 = 1$,$V = 80$)	94.78%	3.42%	1.79%
3D U-Net + SpCycleGAN + BCE ($\mu_1 = 0$, $\mu_2 = 1$,$V = 80$)	95.07%	2.94%	1.99%
3D U-Net + SpCycleGAN + DICE ($\mu_1 = 1$, $\mu_2 = 0$,$V = 80$)	94.76%	3.00%	2.24%
3D U-Net +SpCycleGAN + DICE and BCE ($\mu_1 = 1$, $\mu_2 = 10$,$V = 80$)	95.44%	2.79%	1.76%
3D U-Net +SpCycleGAN + DICE and BCE ($\mu_1 = 1$, $\mu_2 = 10$,$V = 1600$)	95.37%	2.77%	1.86%
3D U-Net +SpCycleGAN + DICE and BCE + PP ($\mu_1 = 1$, $\mu_2 = 10$,$V = 1600$) (Proposed method)	95.56%	2.57%	1.86%

Table 4.3.: Accuracy, Type-I and Type-II errors for known methods and our method on subvolume 2 of Data-I

	Subvolume 2		
Method	Accuracy	Type-I	Type-II
Method [62]	79.25%	20.71%	0.04%
Method [108]	86.78%	13.12%	0.10%
Method [69,70]	88.26%	11.67%	0.07%
Method [43]	92.32%	6.81%	0.87%
3D Encoder-Decoder + CycleGAN + BCE $(\mu_1 = 0,\ \mu_2 = 1, V = 80)$	91.30%	5.64%	3.06%
3D Encoder-Decoder + SpCycleGAN + BCE $(\mu_1 = 0,\ \mu_2 = 1, V = 80)$	92.45%	6.62%	0.92%
3D U-Net + SpCycleGAN + BCE $(\mu_1 = 0,\ \mu_2 = 1, V = 80)$	93.01%	6.27%	0.72%
3D U-Net + SpCycleGAN + DICE $(\mu_1 = 1,\ \mu_2 = 0, V = 80)$	93.03%	6.03%	0.95%
3D U-Net +SpCycleGAN + DICE and BCE $(\mu_1 = 1,\ \mu_2 = 10, V = 80)$	93.63%	5.73%	0.64%
3D U-Net +SpCycleGAN + DICE and BCE $(\mu_1 = 1,\ \mu_2 = 10, V = 1600)$	93.63%	5.69%	0.68%
3D U-Net +SpCycleGAN + DICE and BCE + PP $(\mu_1 = 1,\ \mu_2 = 10, V = 1600)$ (Proposed method)	93.67%	5.65%	0.68%



(a) (b) (c)

(d) (e) (f)

(g) (h)

Fig. 4.14.: 3D visualization of subvolume 1 of Data-I using Voxx [106] (a) original volume, (b) 3D ground truth volume, (c) 3D active surfaces from [62], (d) 3D active surfaces with inhomogeneity correction from [108], (e) 3D Squassh from [69, 70], (f) 3D encoder-decoder architecture from [43], (g) 3D encoder-decoder architecture with CycleGAN, (h) 3D U-Net architecture with SpCycleGAN (Proposed method)

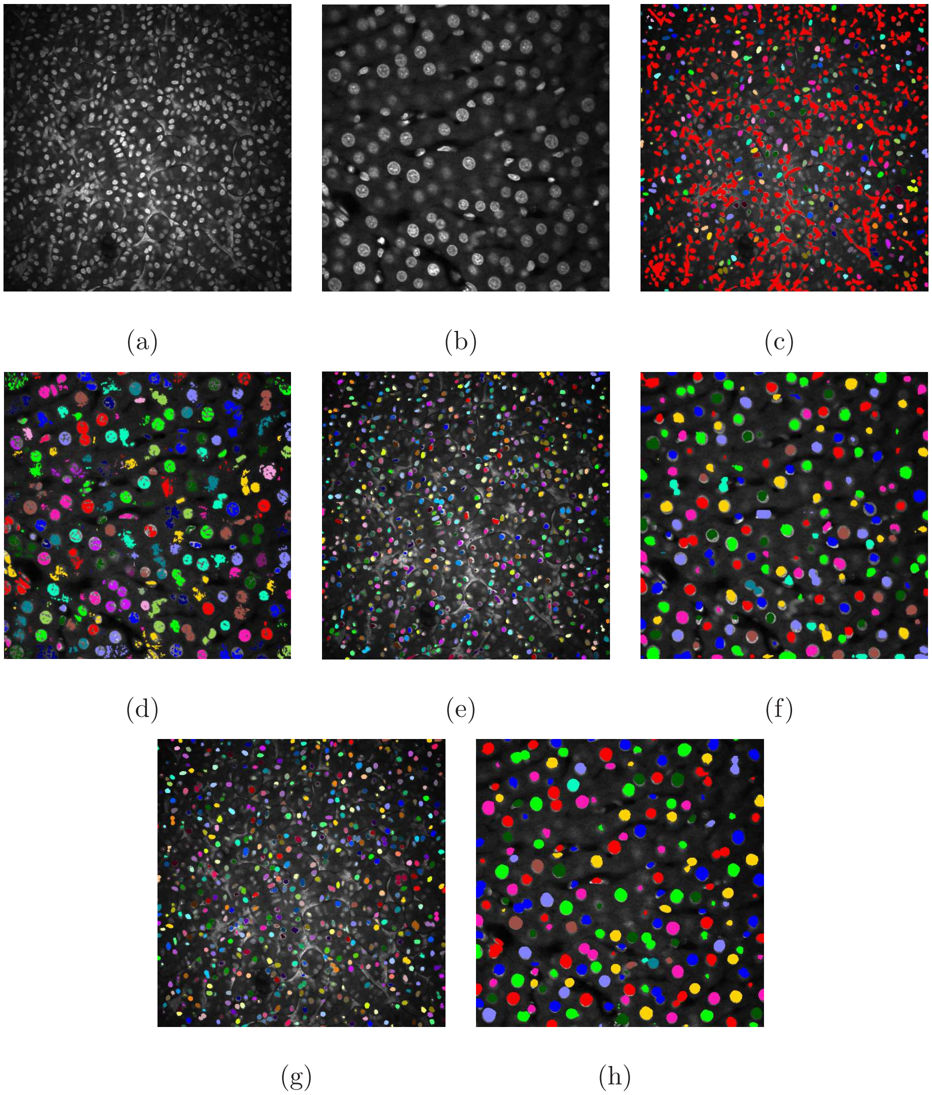


Fig. 4.15.: Original images and their color coded segmentation results of Data-I and Data-II (a) Data-I I_{z66}^{orig}, (b) Data-II I_{z31}^{orig}, (c) Data-I I_{z66}^{seg} using [43], (d) Data-II I_{z31}^{seg} using [43], (e) Data-I I_{z66}^{seg} using 3D encoder-decoder architecture with CycleGAN, (f) Data-II I_{z31}^{seg} using 3D encoder-decoder architecture with CycleGAN, (g) Data-I I_{z66}^{seg} using 3D U-Net architecture with SpCycleGAN (Proposed method), (h) Data-II I_{z31}^{seg} using 3D U-Net architecture with SpCycleGAN (Proposed method)

Table 4.4.: Accuracy, Type-I and Type-II errors for known methods and our method on subvolume 3 of Data-I

	Subvolume 3		
Method	Accuracy	Type-I	Type-II
Method [62]	76.44%	23.55%	0.01%
Method [108]	83.47%	16.53%	0.00%
Method [69, 70]	87.29%	12.61%	0.10%
Method [43]	94.26%	5.19%	0.55%
3D Encoder-Decoder + CycleGAN + BCE ($\mu_1 = 0$, $\mu_2 = 1$,$V = 80$)	94.17%	3.96%	1.88%
3D Encoder-Decoder + SpCycleGAN + BCE ($\mu_1 = 0$, $\mu_2 = 1$,$V = 80$)	93.57%	6.10%	0.33%
3D U-Net + SpCycleGAN + BCE ($\mu_1 = 0$, $\mu_2 = 1$,$V = 80$)	94.04%	5.84%	0.11%
3D U-Net + SpCycleGAN + DICE ($\mu_1 = 1$, $\mu_2 = 0$,$V = 80$)	94.30%	5.22%	0.40%
3D U-Net +SpCycleGAN + DICE and BCE ($\mu_1 = 1$, $\mu_2 = 10$,$V = 80$)	93.90%	5.92%	0.18%
3D U-Net +SpCycleGAN + DICE and BCE ($\mu_1 = 1$, $\mu_2 = 10$,$V = 1600$)	94.37%	5.27%	0.36%
3D U-Net +SpCycleGAN + DICE and BCE + PP ($\mu_1 = 1$, $\mu_2 = 10$,$V = 1600$) (Proposed method)	94.54%	5.10%	0.36%

All segmentation results were evaluated quantitatively based on voxel accuracy, Type-I error and Type-II error metrics, using 3D hand segmented volumes. Here, accuracy $= \frac{n_{\text{TP}}+n_{\text{TN}}}{n_{\text{total}}}$, Type-I error $= \frac{n_{\text{FP}}}{n_{\text{total}}}$, Type-II error $= \frac{n_{\text{FN}}}{n_{\text{total}}}$, where n_{TP}, n_{TN}, n_{FP}, n_{FN}, n_{total} are defined to be the number of true-positives (voxels segmented as nuclei correctly), true-negatives (voxels segmented as background correctly), false-positives (voxels falsely segmented as nuclei), false-negatives (voxels falsely segmented as background), and the total number of voxels in a volume, respectively.

The quantitatively evaluations for the subvolumes are shown in Table 4.2, 4.3 and 4.4. Our proposed method outperforms other compared methods. The smaller Type-I error shows our proposed method successfully rejects non-nuclei structures during segmentation. Also, our proposed method has reasonably low Type-II errors compared to other segmentation methods. Moreover, in this table, we show that our proposed SpCycleGAN creates better paired synthetic volumes which reflects in segmentation accuracy. Instead of 3D encoder-decoder structure, we use 3D U-Net which leads to better results since 3D U-Net has skip connections that can preserve spatial information. In addition, the combination of two loss functions such as the Dice loss and the BCE loss turns out to be better for the segmentation task in our application. In particular, the Dice loss constrains the shape of the nuclei segmentation whereas the BCE loss regulates voxelwise binary prediction. It is observed that training with more synthetic volumes can generalize our method to achieve better segmentation accuracy. Finally, the postprocessing (PP) that eliminates small components helps to improve segmentation performance.

To make this clear, segmentation results were color coded using 3D connected component labeling and overlaid on the original volumes. The method from [43] cannot distinguish between nuclei and non-nuclei structures including noise. This is especially recognizable from segmentation results of Data-I in which multiple nuclei and non-nuclei structures are colored with the same color. As can be observed from Figure 4.15(e) and 4.15(f), segmentation masks are smaller than nuclei size and suffered from location shifts. Conversely, our proposed method shown in Figure 4.15(g) and 4.15(h) segments nuclei with the right shape at the correct locations.

4.4 MTU-Net [2]

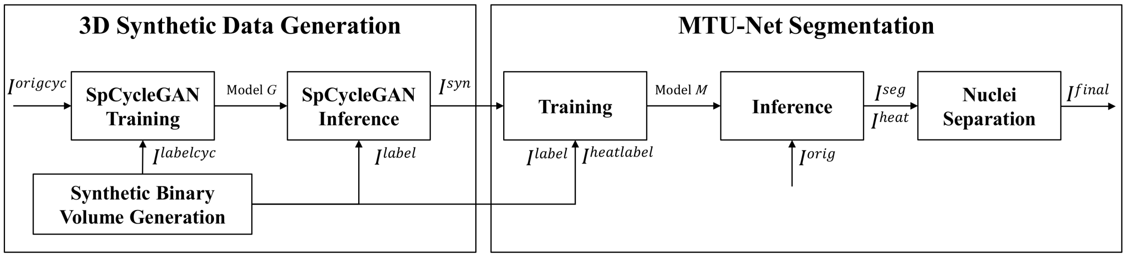


Fig. 4.16.: Block diagram of our method

Figure 4.16 shows a block diagram of our method. We denote I as a 3D image volume of size $X \times Y \times Z$. Note that I_{z_p} is a p^{th} focal plane image, of size $X \times Y$, along the z-direction in a volume, where $p \in \{1, \ldots, Z\}$. In addition, let $I_{\left(q_i:q_f,r_i:r_f,p_i:p_f\right)}$ be a subvolume of I, whose x-coordinate is $q_i \leq x \leq q_f$, y-coordinate is $r_i \leq y \leq r_f$, z-coordinate is $p_i \leq z \leq p_f$, where $q_i, q_f \in \{1, \ldots, X\}$, $r_i, r_f \in \{1, \ldots, Y\}$, $p_i, p_f \in \{1, \ldots, Z\}$, $q_i \leq q_f$, $r_i \leq r_f$, and $p_i \leq p_f$. For example, $I^{seg}_{(241:272,241:272,131:162)}$ is a subvolume of a segmented volume, I^{seg}, where the subvolume is cropped between 241st slice and 272nd slice in x-direction, between 241st slice and 272nd slice in y-direction, and between 131st slice and 162nd slice in z-direction.

As shown in Figure 4.16, our proposed method is a two-stage method that consists of synthetic volume generation and MTU-Net segmentation. We first train a spatially constrained CycleGAN (SpCycleGAN) with synthetic binary volumes, $I^{labelcyc}$, and a subvolume of the original image volumes, $I^{origcyc}$ to obtain a generative model denoted as model G. To create MTU-Net training volumes, a new set of synthetic binary volume, I^{label}, and its corresponding heat map, $I^{heatlabel}$, are generated. A set of synthetic microscopy volumes I^{syn} are generated using model G with I^{label}. Note that I^{label} is a binary segmentation mask whereas $I^{heatlabel}$ indicates the centroids of nuclei. Here, I^{label} and $I^{heatlabel}$ serve as the segmentation labels and heat map labels of I^{syn}. A multi-task network, MTU-Net, is trained with I^{syn}, I^{label} and $I^{heatlabel}$ to

[2] This is joint work with Ms. Shuo Han and Mr. Soonam Lee

obtain a model M. Also, I^{orig} is the original fluorescence microscopy volume. The corresponding segmented volume, I^{seg}, and heat map, I^{heat} of I^{orig} can be obtained using model M on I^{orig}. Finally, a nuclei separation method marker-controlled watershed [65] is used to separate overlapped nuclei in I^{seg}. This produces the final segmentation I^{final} of I^{orig}.

4.4.1 3D Convolutional Neural Network

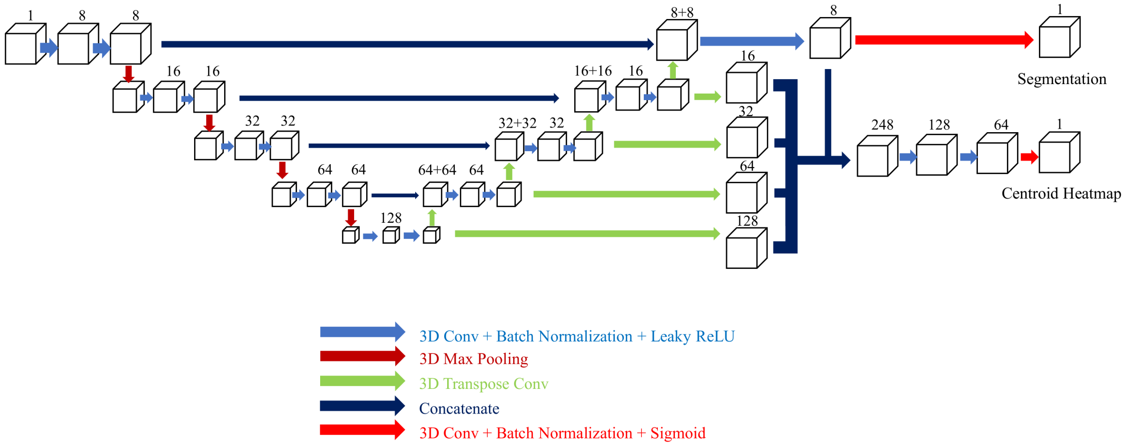


Fig. 4.17.: Architecture of our MTU-Net

Figure 4.17 shows architecture of our network. Our network is a multi-task U-Net that outputs a 3D heatmap of the location of the nuclei and a probability map of binary volumetric segmentation. The 3D heatmap is used to separate overlapped nuclei in the binary volumetric segmentation and the detail is described in 4.4.2. After separating overlapped nuclei, our method is able to produce instance segmentation of the nuclei. The binary segmentation branch is the same as described in our previous work [53]. Additionally, we extract spatial information of each layer of the decoder and concatenate them together to form a branch that estimates the 3D heatmap of the nuclei. A mean squared error is used to measure the difference between the predicted 3D heatmap and the label of the 3D heatmap while a combination of the Dice loss and binary cross-entropy loss is used to measure the difference between the predicted binary volumetric segmentation and the label of the segmentation. Therefore, the total training loss of our network can be expressed as a linear combination of the Dice loss, the binary cross entropy loss and the mean squared error such that

$$\begin{aligned}\mathcal{L}_{seg}(T, S, C, D) = \mu_1 \mathcal{L}_{Dice}(T, S) + \mu_2 \mathcal{L}_{BCE}(T, S) \\ + \mu_3 \mathcal{L}_{MSE}(C, D) \qquad (4.4)\end{aligned}$$

where

$$\begin{aligned}\mathcal{L}_{Dice}(T, S) &= \frac{2(\sum_{i=1}^{P} t_i s_i)}{\sum_{i=1}^{P} t_i^2 + \sum_{i=1}^{P} s_i^2} \\ \mathcal{L}_{BCE}(T, S) &= -\frac{1}{P} \sum_{i=1}^{P} t_i \log(s_i) + (1 - t_i) \log(1 - s_i) \\ \mathcal{L}_{MSE}(C, D) &= \frac{1}{P} \sum_{i=1}^{P} (c_i - d_i)^2,\end{aligned}$$

respectively [101]. Note that T is the set of the groundtruth values of volumetric binary segmentation while S is a prediction of binary volumetric segmentation. $t_i \in T$ and $s_i \in S$ are a groundtruth value at i^{th} voxel location and a value of prediction at i^{th} voxel location. Also, C is the set of groundtruth values of the 3D heatmap and $c_i \in C$ is a groundtruth value of the 3D heatmap at i^{th} voxel location. Similarly, D is a predicted 3D heatmap and $d_i \in D$ is a value of the predicted 3D heatmap at i^{th} voxel location. Lastly, P is the number of entire voxels and μ_1, μ_2, and μ_3 serve as the weight coefficient between to loss terms in Equation (4.4). Our proposed network produces an volumetric binary segmentation and 3D heatmap with the same size of a input grayscale volume of size of $64 \times 64 \times 64$. To train our model M, V pairs of I^{syn}, I^{label}, and $I^{heatlabel}$ are used.

For the inference of our network, a moving inference window with size of $64 \times 64 \times 64$ is slided through the entire volume starting from top to bottom and from left to right. First, a symmetric padding was peformed to pad the original volume I^{orig} by 16 voxels in x, y and z-direction. Since partial included nuclei structures may create artifacts near the boundaries of the moving window, the stride of the sliding window was set to 32 in x, y and z-directions and only the segmentation of size of $32 \times 32 \times 32$ at the center of the window is used to generate the corresponding subvolume of I^{seg}. More details were described in our previous work [43].

4.4.2 Nuclei Separation

To achieve instance segmentation, our network uses the 3D heatmap with the binary volumetric segmentation. An additional nuclei separation step is employed on binary volumetric segmentation to separate overlapped nuclei. Here, we describe two different approaches to separate overlapped nuclei such as quasi 3D watershed and marker controlled watershed.

Quasi 3D Watershed

Our previous method [53] achieves promising segmentation results in terms of voxel accuracy but fail to identify overlapped nuclei. We use watershed which is a well-known and widely used technique to solve this problem. Since our goal is to produce a volumetric segmentation, a 3D watershed is prefered. However 3D watershed is computationally expensive when the input volume is large. Instead of using a 3D, A 2D watershed [63] is used on the 3D segmentation in three different direction sequentially to separate overlapped nuclei in a quasi 3D manner.

Marker Controlled Watershed

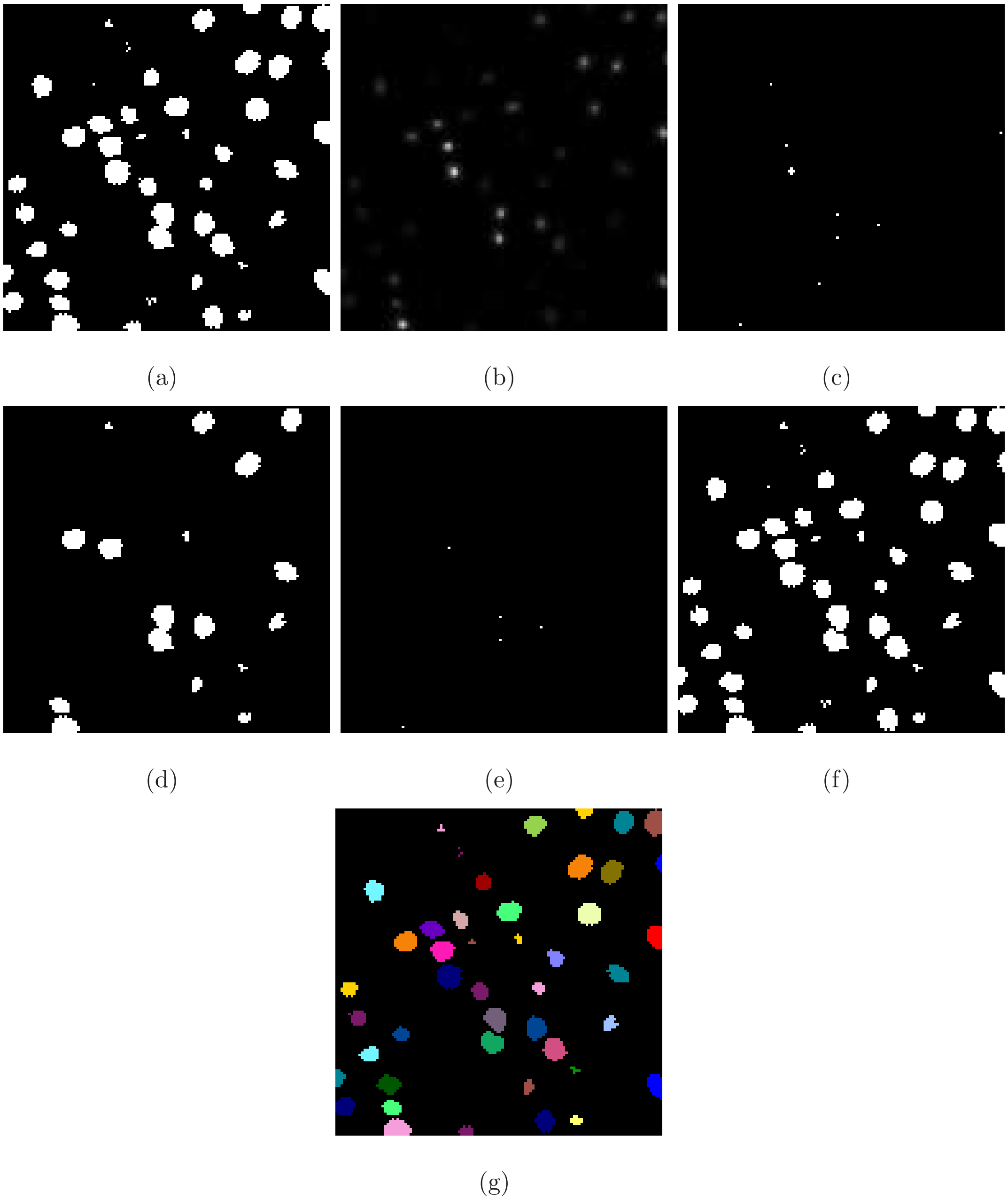
Watershed algoritm tends to oversegment objects into multiple small pieces. Here, a marker controlled watershed is used to minimize oversegmentation problems in the nuclei separation [65]. First, a non-maximum suppression is used on the heatmap followed by a 3D connected components analysis to extract the centroids of the nuclei, I^{ct}. More specifically, the non-maximum supression uses a ball shape sliding window with radius of R. R is selected according to the real size of the nuclei. Then, I^{ct} is dilated to a ball with radius of $\frac{R}{3}$. In order to reduce over segmentation of watershed technique, we only use marker controlled watershed on the components in I^{seg} that contain no less than two centroids in I^{ct}. A marker map $I^{markseg}$ can be generated by finding the centroids objects in I^{seg} that contain no less than two centroids in I^{ct}. I^{markct} are generated by finding the centroids that overlapped with $I^{markseg}$. We use [65] to separate overlapped nuclei from $I^{markseg}$ according to marker map I^{markct}. The final segmentation I^{final} is obtained by adding the output of marker controlled watershed. The sample results of different stages of our proposed method are shown in Figure 4.18.

4.4.3 Experimental Results

We tested our proposed method on four different rat kidney data sets and one rat cardiomyocytes data set. Data-I contains grayscale images with size X = 512 × Y = 512 × Z = 512. Data-II contains grayscale images with size X = 512 × Y = 512 × Z = 415. Data-III contains grayscale images with size X = 512 × Y = 512 × Z = 32. Data-IV contains grayscale images with size X = 512 × Y = 512 × Z = 300. Data-V contains grayscale images with size X = 512 × Y = 512 × Z = 157. Note that Data-I, II, III, and V are obtained from rat kidney whereas Data-IV is obtained from rat cardiomyocytes.

Synthetic Generation

Our SpCycleGAN is implemented in PyTorch using the Adam optimizer with constant learning of 0.0002 in the first 100 epochs and gradually decayed learning rate from 0.0002 to 0 in the second 100 epochs. We use Resnet 9 blocks for both network G, F and H. For each of the data, the sizes of $I^{labelcyc}$ and $I^{origcyc}$ were both 128 × 128 × 128. Here, $I^{origcyc}$ is a subvolume of I^{orig}. A 64 × 64 2D random cropping was used to augment training images before training. For Data-I, Data-II, and Data-III, SpCycleGAN generative models G^{Data-I},$G^{Data-II}$ and $G^{Data-III}$ were



(a) (b) (c)

(d) (e) (f)

(g)

Fig. 4.18.: Sample results of different stages of our proposed method. (a) I^{seg} (b) I^{heat} (c) dilated I^{ct} (d) $I^{markseg}$ (e) I^{markct} (f) I^{final} (g) color result

trained individually using $\lambda_1 = \lambda_2 = 10$. For Data-IV, SpCycleGAN generative model $G^{Data-IV}$ was trained using $\lambda_1 = \lambda_2 = 50$ to penalize more on spatial constrains since Data-IV contains more directional pattern. For each data, 80 sets of I^{syn}, I^{label} and

$I^{heatlabel}$ were generated using its own generative model. The size of each volume of I^{syn}, I^{label} and $I^{heatlabel}$ is 64 × 64 × 64.

MTU-Net Segmentation

Table 4.5.: True positive, False positive, False negative, Precision, Recall and F1 Scores for known methods and our method on Data-I

	Data-I					
Method	N_{TP}	N_{FP}	N_{FN}	P	R	$F1$
Otsu [55] + Quasi 3D watershed	151	22	132	87.28%	53.36%	66.23%
CellProfiler [109]	59	14	223	80.82%	20.92%	33.24%
Squassh [69, 70]	109	12	174	90.08%	38.52%	53.96%
Method [53]	228	22	50	91.20%	82.01%	86.36%
Method [53] + Quasi 3D watershed	261	31	13	89.38%	95.26%	92.23%
MTU-Net (Proposed)	260	20	17	92.86%	93.86%	93.36%

Table 4.6.: Voxel Accuracy, Type-I and Type-II for known methods and our method on Data-I

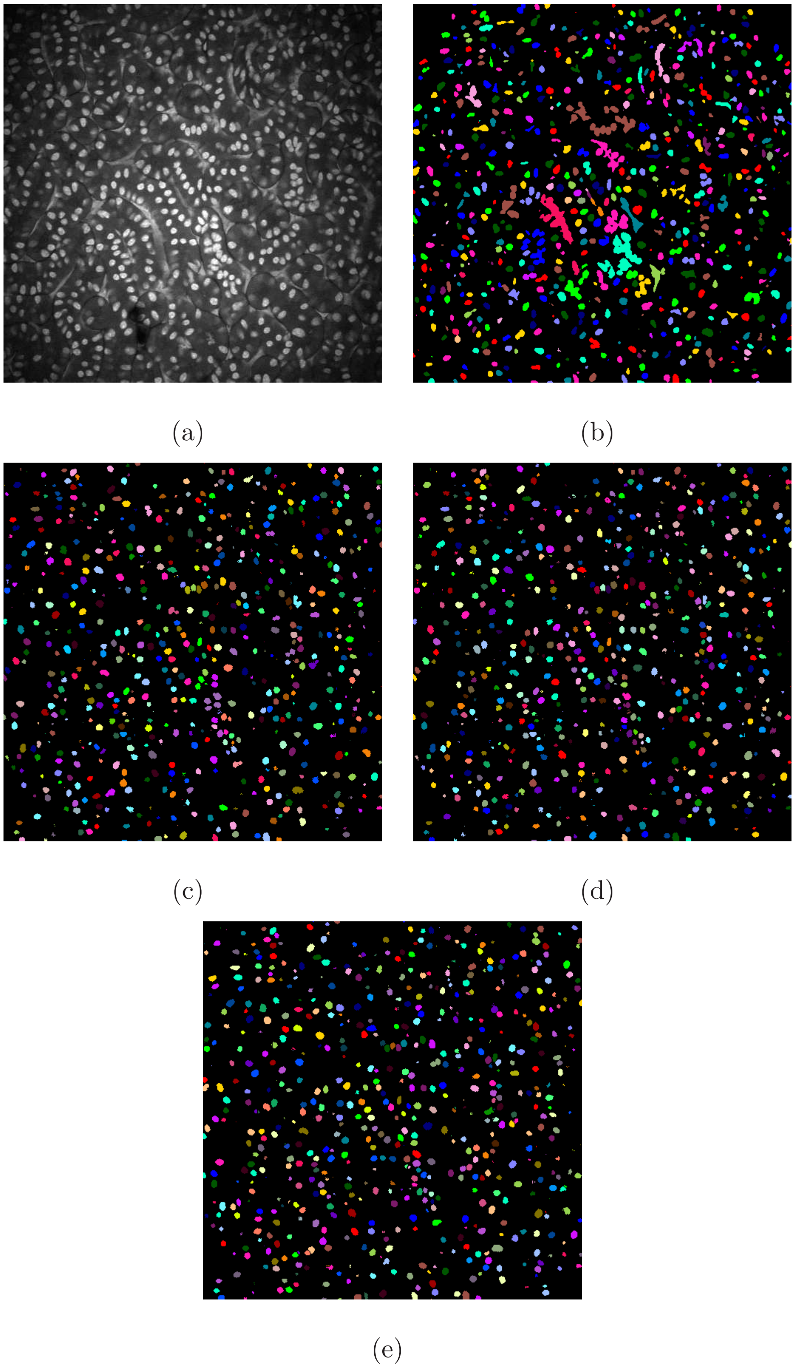
	Data-I		
Method	Voxel Accuracy	Type-I	Type-II
Otsu [55] + Quasi 3D watershed	81.89%	17.88%	0.23%
CellProfiler [109]	78.02%	21.67%	0.31%
Squassh [69, 70]	86.48%	11.87%	1.65%
Method [53]	95.68%	1.33%	2.99%
Method [53] + Quasi 3D watershed	95.73%	1.49%	2.78%
MTU-Net (Proposed)	95.68%	1.86%	2.46%

Table 4.7.: True positive, False positive, False negative, Precision, Recall and F1 Scores for known methods and our method on Data-III

	Data-III					
Method	N_{TP}	N_{FP}	N_{FN}	P	R	$F1$
Otsu [55] + Quasi 3D watershed	223	47	69	82.59%	76.37%	79.36%
CellProfiler [109]	218	37	78	85.49%	73.65%	79.13%
Squassh [69, 70]	243	22	79	91.70%	75.47%	82.79%
Method [53]	321	92	3	92.18%	83.38%	87.56%
Method [53] + Quasi 3D watershed	317	47	5	87.09%	98.45%	92.42%
MTU-Net (Proposed)	303	30	18	91.27%	94.41%	92.82%

Table 4.8.: Voxel Accuracy, Type-I, and Type-II for known methods and our method on Data-III

	Data-III		
Method	Voxel Accuracy	Type-I	Type-II
Otsu [55] + Quasi 3D watershed	93.95%	2.53%	3.51%
CellProfiler [109]	93.95%	2.66%	3.39%
Squassh [69, 70]	94.84%	4.46%	0.70%
Method [53]	92.19%	1.93%	5.88%
Method [53] + Quasi 3D watershed	92.29%	1.79%	5.92%
MTU-Net (Proposed)	92.69%	1.41%	5.90%



(a)

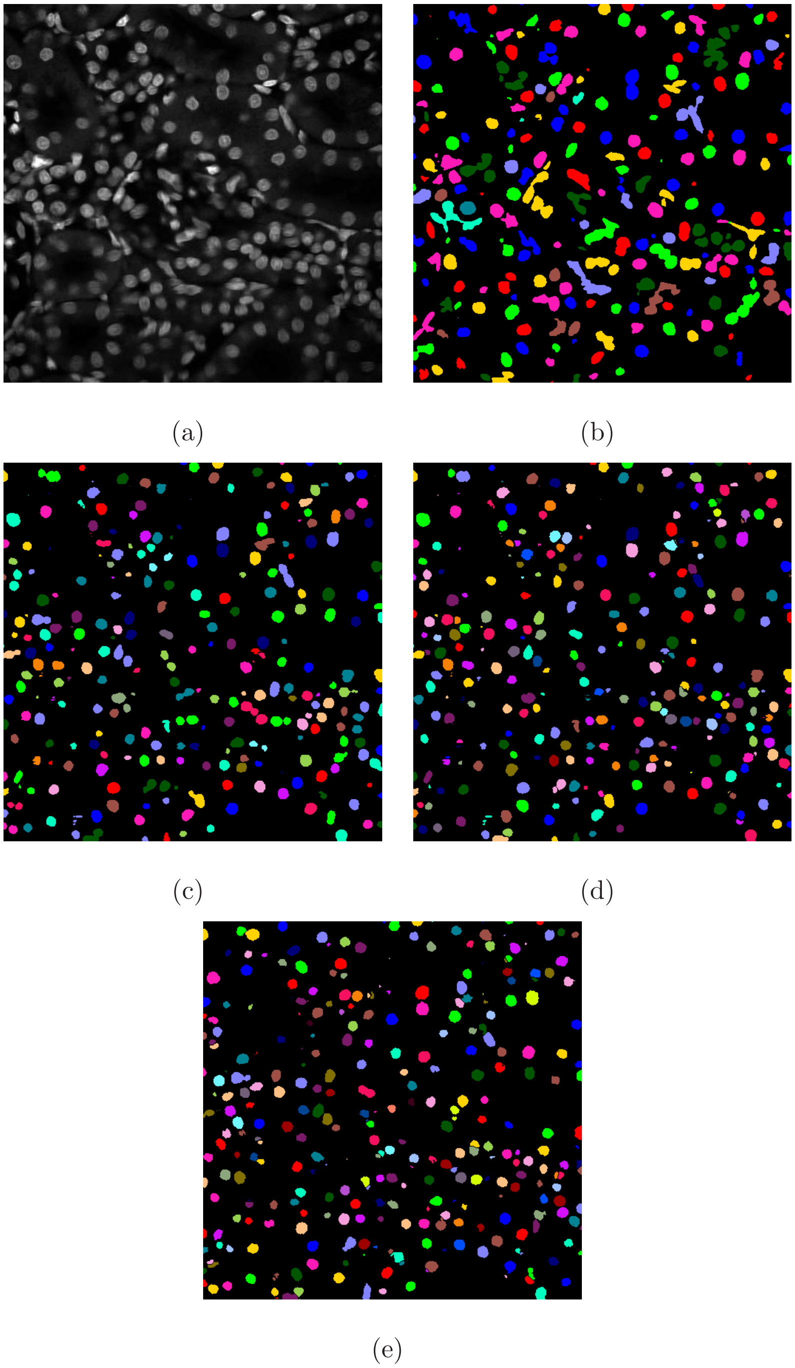
(b)

(c)

(d)

(e)

Fig. 4.19.: Sample results of Data-I (a) Original microscopy images (b) Segmentations of Squassh (c) Segmentations of method [53] (d) Segmentations of method [53] + Quasi 3D watershed (e) Segmentations of MTU-Net


(a)

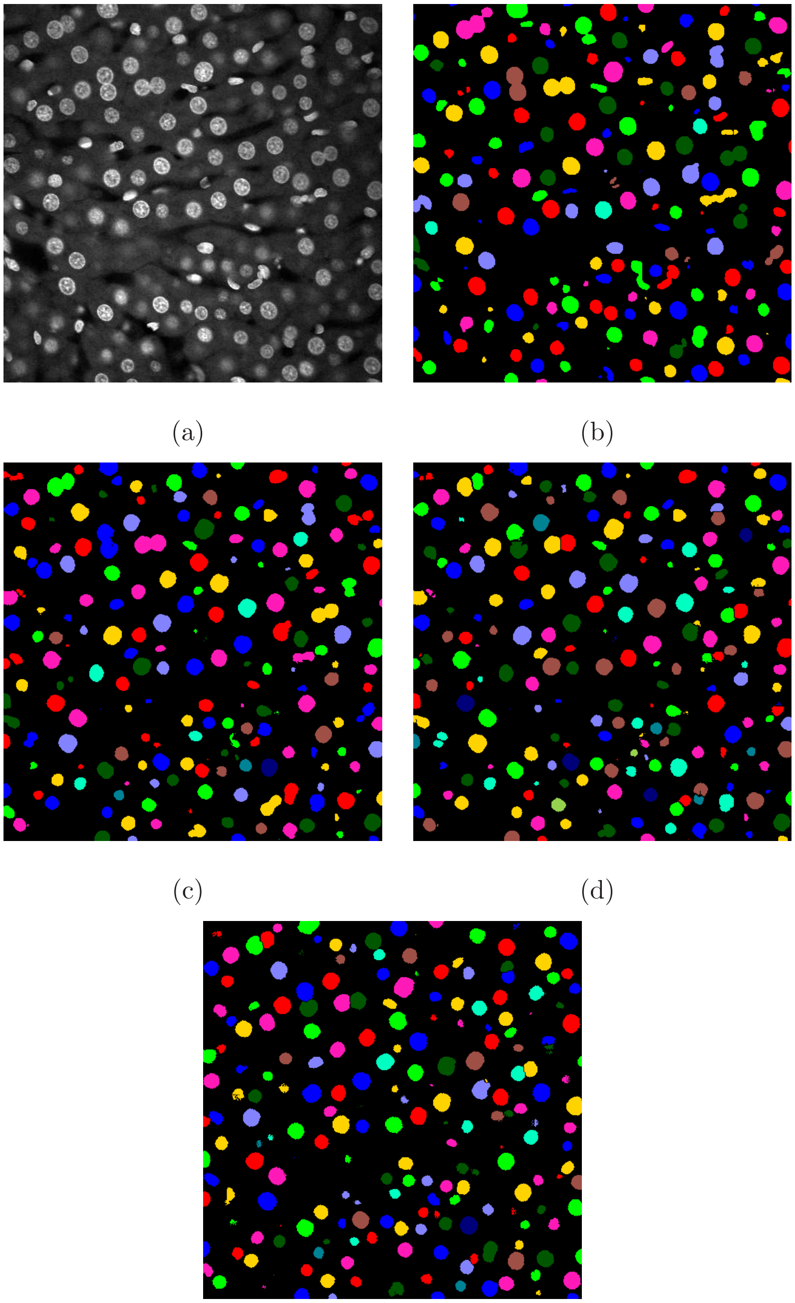
(b)

(c)

(d)

(e)

Fig. 4.20.: Sample results of Data-II (a) Original microscopy images (b) Segmentations of Squassh (c) Segmentations of method [53] (d) Segmentations of method [53] + Quasi 3D watershed (e) Segmentations of MTU-Net



(a)

(b)

(c)

(d)

(e)

Fig. 4.21.: Sample results of Data-III (a) Original microscopy images (b) Segmentations of Squassh (c) Segmentations of method [53] (d) Segmentations of method [53] + Quasi 3D watershed (e) Segmentations of MTU-Net

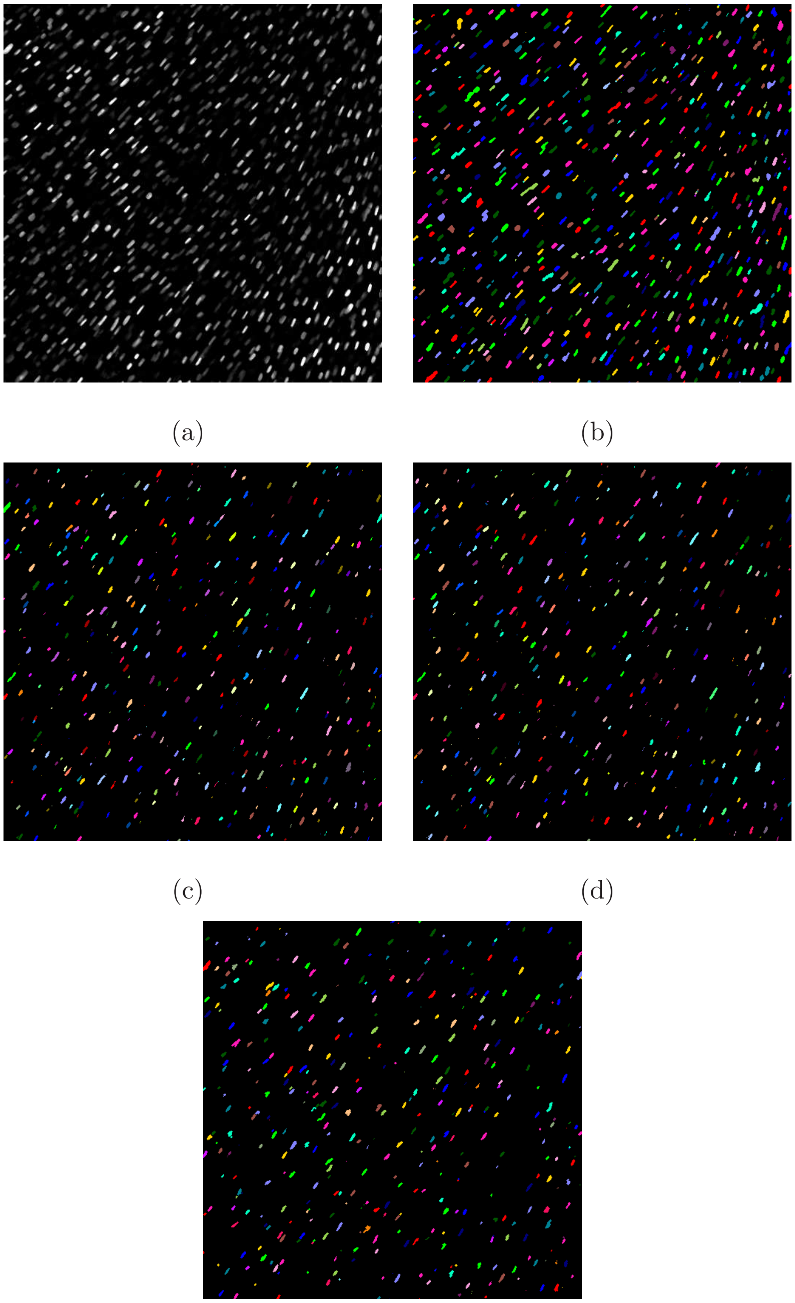


Fig. 4.22.: Sample results of Data-IV (a) Original microscopy images (b) Segmentations of Squassh (c) Segmentations of method [53] (d) Segmentations of method [53] + Quasi 3D watershed (e) Segmentations of MTU-Net

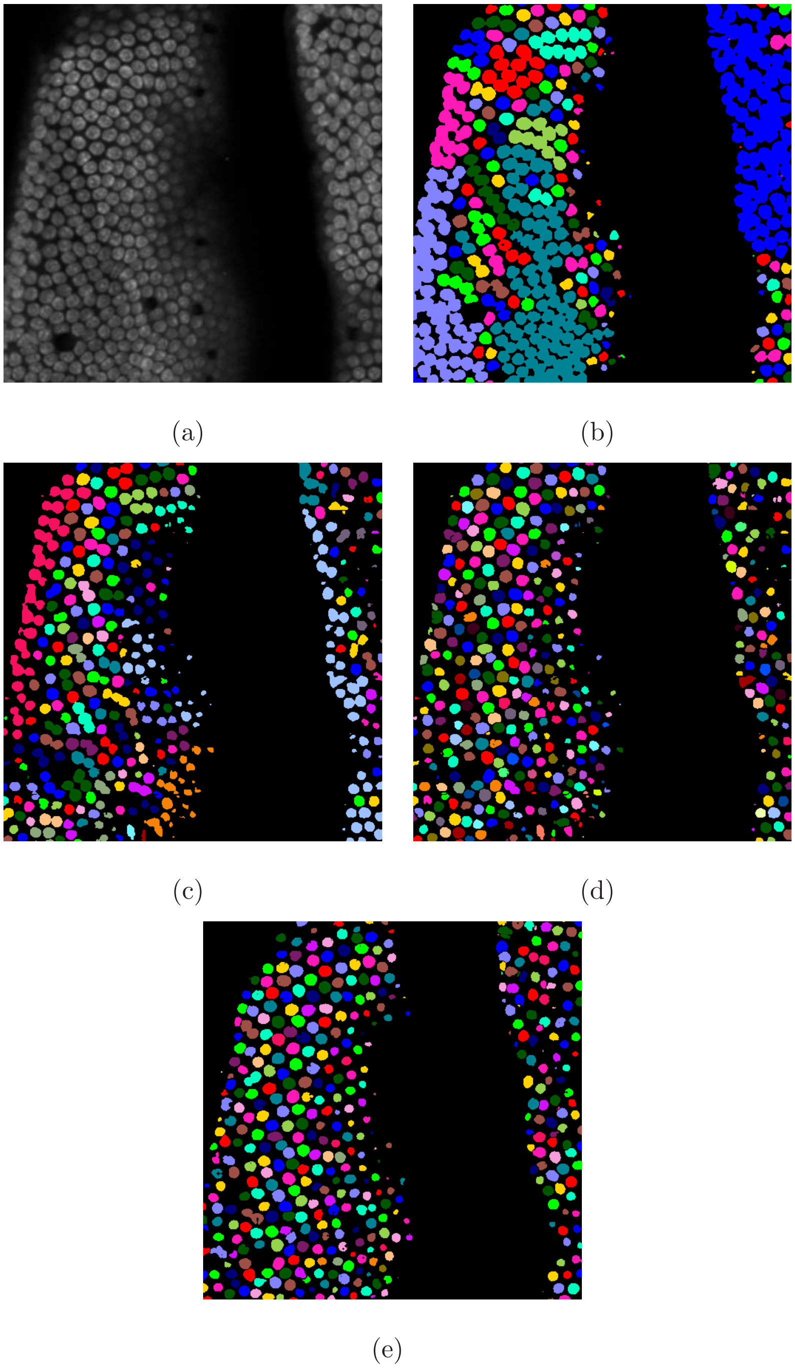
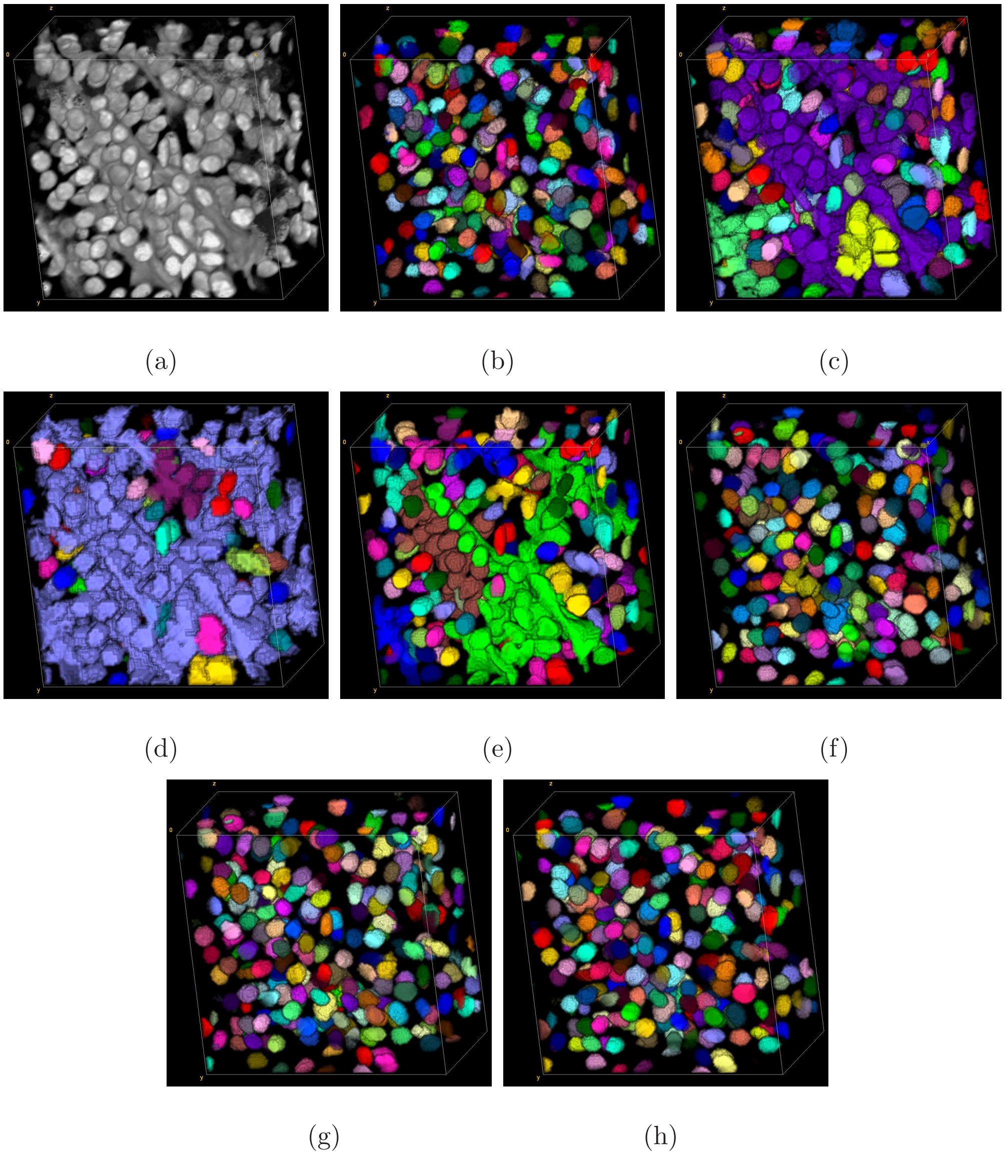


Fig. 4.23.: Sample results of Data-V (a) Original microscopy images (b) Segmentations of Squassh (c) Segmentations of method [53] (d) Segmentations of method [53] + Quasi 3D watershed (e) Segmentations of MTU-Net



(a) (b) (c)

(d) (e) (f)

(g) (h)

Fig. 4.24.: 3D visualization of different methods of subvolume of Data-I. (a) Original volume (b) Groundtruth volume (c) Otsu + Quasi 3D watershed (d) CellProfiler (e) Squassh (f) Method [110] (g) Method [110] + Quasi 3D watershed (h) MTU-Net (Proposed)

Our MTU-Net is also implemented in PyTorch using Adam optimizer with learning rate of 0.001. For each of the Data-I, Data-II, Data-III and Data-IV, MTU-Net models M^{Data-I},$M^{Data-II}$,$M^{Data-III}$ and $M^{Data-IV}$ were trained individually with

80 sets of $I^{syn}, I^{label}, I^{heatmap}$. The weights of MTU-Net loss function were used as $\mu_1 = 1$ and $\mu_2 = \mu_3 = 10$. We tested Data-I, Data-II, Data-III and Data-IV with model M^{Data-I}, $M^{Data-II}$, $M^{Data-III}$, and $M^{Data-IV}$ respectively. Additionally, we tested Data-V with the model $M^{Data-II}$ since they shares similar characteristic of nuclei. For nuclei separation step, we $R^{Data-I} = 5$, $R^{Data-II} = 7$, $R^{Data-III} = 13$, $R^{Data-IV} = 5$, and $R^{Data-V} = 6$. For the convenience of visualization, we used 3D connected components to identify individual nuclei and assigned them with different color. Small 3D connected components that less than 20 voxels are removed at the end.

We evaluate our segmentation on Data-I and Data-III. Two groundtruth volumes, $I^{gt,Data-I}$ and $I^{gt,Data-III}$, are manually anotated using ITK-SNAP [111]. $I^{gt,Data-I}$ is 128 × 128 × 64 and corresponds to $I^{orig,Data-I}_{(193:320,193,320,31:94)}$. $I^{gt,Data-III}$ is 512 × 512 × 32 and corresponds to the entire $I^{orig,Data-III}$. To evaluate the segmentation, both voxel-based evaluation and object-based evaluation are used. For voxel-based evaluation, Type-I and Type-II error metric was used. $\text{voxel accuracy} = \frac{n_{\text{TP}}+n_{\text{TN}}}{n_{\text{total}}}$, $\text{Type-I error} = \frac{n_{\text{FP}}}{n_{\text{total}}}$, $\text{Type-II error} = \frac{n_{\text{FN}}}{n_{\text{total}}}$, where n_{TP}, n_{TN}, n_{FP}, n_{FN}, n_{total} are defined to be the number of true-positives (voxels segmented as nuclei correctly), true-negatives (voxels segmented as background correctly), false-positives (voxels falsely segmented as nuclei), false-negatives (voxels falsely segmented as background), and the total number of voxels in a volume, respectively. For object-based evaluation, F1 score (F1), Precision (P) and Recall (R) [112, 113] were obtained as:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}}, R = \frac{N_{TP}}{N_{TP} + N_{FN}}, F1 = \frac{2PR}{P + R}, \tag{4.5}$$

where N_{TP} is the number of true-positive, N_{FP} is the number of false-positive, N_{TN} is the number of true-negative,and N_{FN} is the number of false-negative. Here, a true-postive is defined as the segmentation of a nucleus overlap more than 50% with corresponding nucleus in the groundtruth. Otherwise, it a false-positive. A true-negative is defined as the segmentation of a nucleus overlap less than 50% with corresponding nucleus in the groundtruth or no corresponding nucleus presents in the groundtruth.

Our method was compared to 6 different methods including Otsu [55] + quasi 3D watershed, CellProfiler [109], Squassh [69, 70], our previous work [53], and [53] + quasi 3D watershed. Otsu cannot separate overlapped nuclei so the quasi 3D watershed was used on the results of Otsu. CellProfiler is a cell image analysis tool that are commonly used in biological researches. We used CellProfiler for nuclei segmentation that includes contrast enhancement, median filtering, Otsu thresholding, hole removal, and watershed. For Squassh, the default parameters were used for testing. Our previous work [53] is trained with the same synthetic data that MTU-Net used for training.

The best four of the compared methods of five different data sets are shown in Figure 4.19, 4.20, 4.21, 4.22 and 4.23. As shown in the Figure 4.19, 4.20, 4.21, 4.22 and 4.23, Squassh is able to segment nuclei as individual objects if the original volume is sparse and clear but failed otherwise, especially when non-nuclei structure and noise are presented. Our previous work [53] is able to segment nuclei accurately but not able to separate overlapped nuclei. With a quasi 3D watershed used after [53], the overlapped nuclei are observed to be identified as indivdual nuclei for the most of the situations. However, if multiple nuclei are overlapped with each others, this method may fail to separate the overlapped objects accurately. Our proposed method uses a heatmap of centroids to locate the nuclei in a overlapped objects and uses mark controlled watershed to separate them accurately. We also visualized the results of each methods in 3D using ImageJ Volume Viewer [114]. A comparison of 3D visualization were also shown in 4.24.

In Table 4.5 and 4.7, it was shown that our proposed method reduces the number of false-positive in the object-based evaluation of both of the data sets. It means our proposed method is able to separate nuclei more accurately compared to others. However, due to the limitaion of non-maximum suppression, the increasing number of false-negative is also observed. Our proposed method also achieved high voxel accuracy since our proposed method can segment the shape of the nuclei accurately.

5. DISTRIBUTED AND NETWORKED ANALYSIS OF VOLUMETRIC IMAGE DATA (DINAVID)

5.1 System Overview[1]

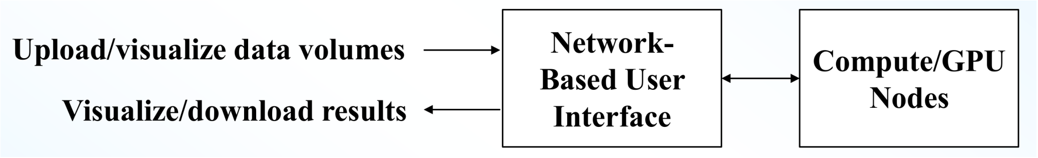
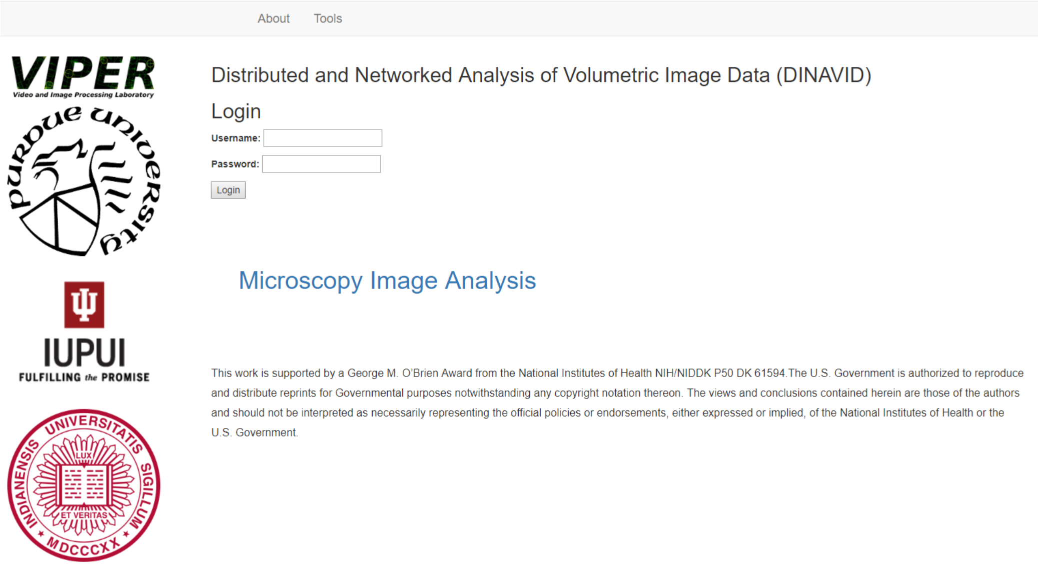

Upload/visualize data volumes
Visualize/download results
Network-Based User Interface
Compute/GPU Nodes

Fig. 5.1.: System diagram of DINAVID

We designed and developed a web-based microscopy image analysis system. We call this system the Distributed and Networked Analysis of Volumetric Image Data (DINAVID). This system is designed for fast and accurate analysis of large scale microscopy volumes. As shown in Figure 5.1, our system consists of web-based user interface and computing clusters that contains high performance GPUs. User will be able to upload and download data using web-based user interface. Also, built in image previewer and built in 3D volume visualization tools are integrated for visualizing data before and after processes.

In Figure 5.2, user will need to login to our system using our issued credential. Currently, we only issue credential upon request. As shown in Figure 5.3, users will see the tutorial of our system and our project information once they login. In the ”Tool” tap, a upload function will allow user to uploaded they data into system using the blue ”Upload Images” button. As shown in Figure 5.4, at the right top of the page, user can delete all the images using the red ”Delete Uploaded Images” button. Currently, our system only support 2D image slices and will support 3D image in the future.

[1] This is joint work with Ms. Shuo Han, Mr. Soonam Lee and Dr. David J. Ho



Distributed and Networked Analysis of Volumetric Image Data (DINAVID)

Login

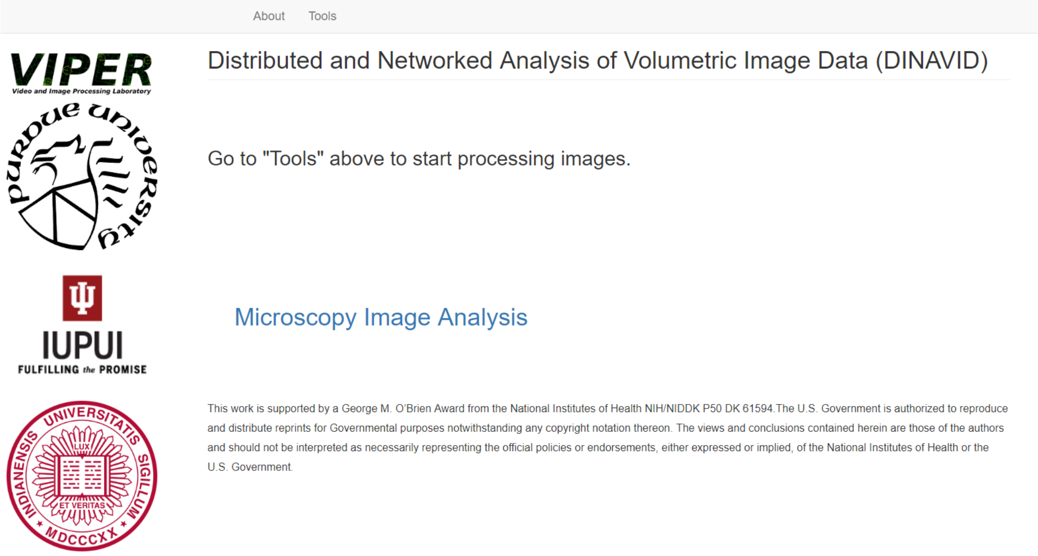
Username:

Password:

Microscopy Image Analysis

This work is supported by a George M. O'Brien Award from the National Institutes of Health NIH/NIDDK P50 DK 61594.The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the National Institutes of Health or the U.S. Government.

Fig. 5.2.: Login page of DINAVID



Distributed and Networked Analysis of Volumetric Image Data (DINAVID)

Go to "Tools" above to start processing images.

Microscopy Image Analysis

This work is supported by a George M. O'Brien Award from the National Institutes of Health NIH/NIDDK P50 DK 61594.The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the National Institutes of Health or the U.S. Government.

Fig. 5.3.: Home page of DINAVID

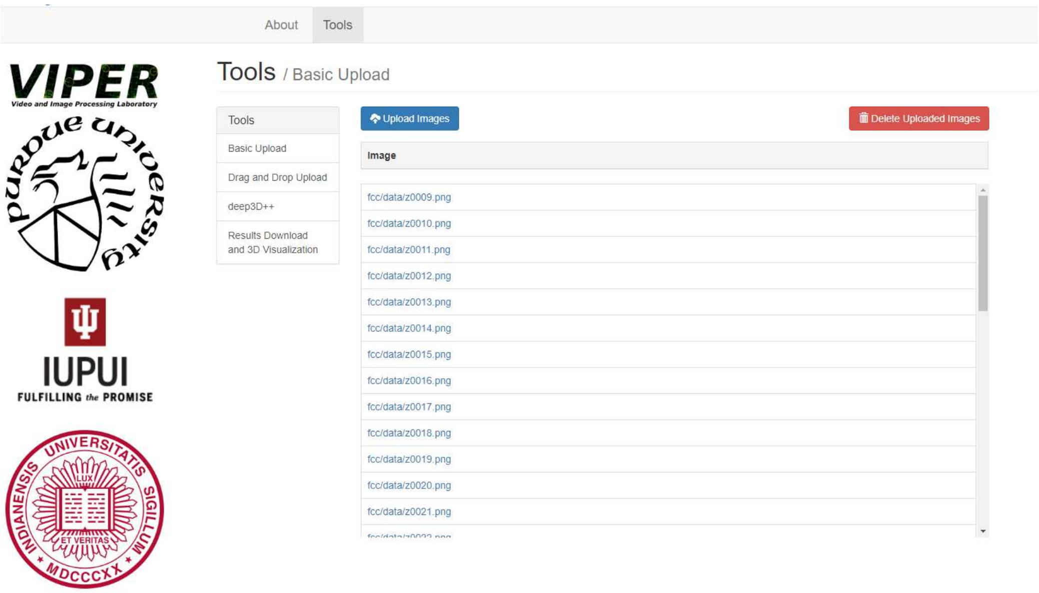


Image
fcc/data/z0009.png
fcc/data/z0010.png
fcc/data/z0011.png
fcc/data/z0012.png
fcc/data/z0013.png
fcc/data/z0014.png
fcc/data/z0015.png
fcc/data/z0016.png
fcc/data/z0017.png
fcc/data/z0018.png
fcc/data/z0019.png
fcc/data/z0020.png
fcc/data/z0021.png

Fig. 5.4.: Data upload page of DINAVID

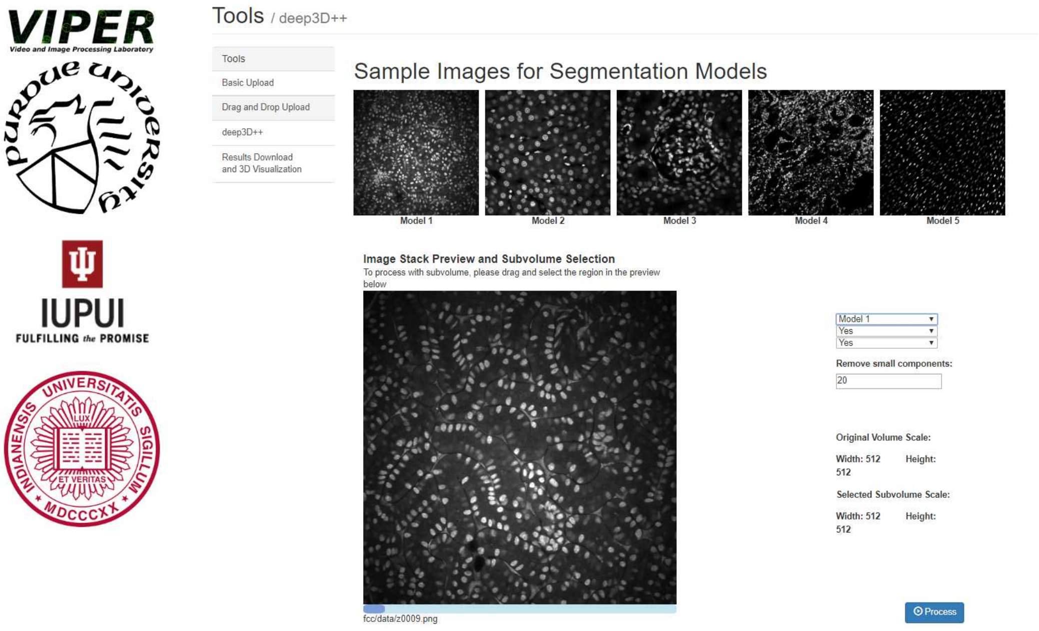


Image Stack Preview and Subvolume Selection

To process with subvolume, please drag and select the region in the preview below

fcc/data/z0009.png

Model 1
Yes
Yes

Remove small components: 20

Original Volume Scale:
Width: 512 Height: 512

Selected Subvolume Scale:
Width: 512 Height: 512

Fig. 5.5.: Segmentation tool page of DINAVID

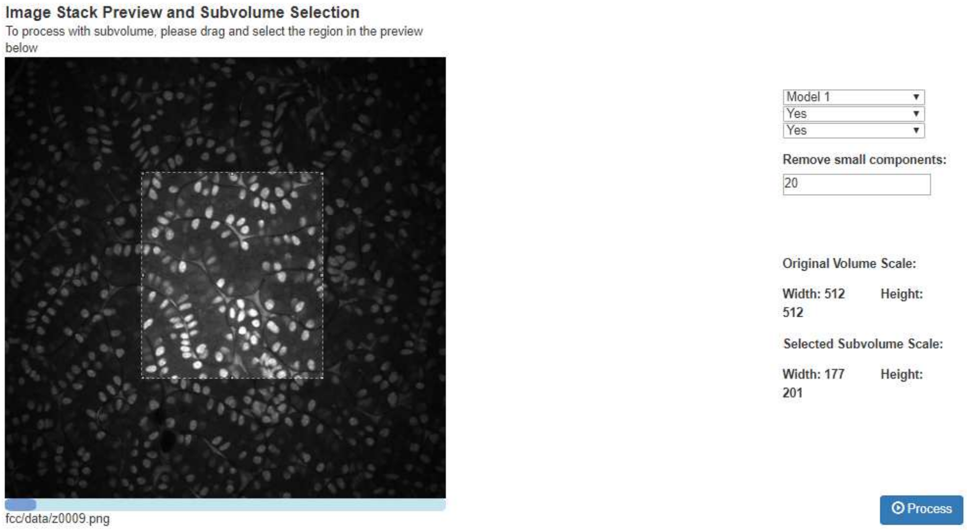


Image Stack Preview and Subvolume Selection

To process with subvolume, please drag and select the region in the preview below

fcc/data/z0009.png

Model 1

Yes

Yes

Remove small components: 20

Original Volume Scale:

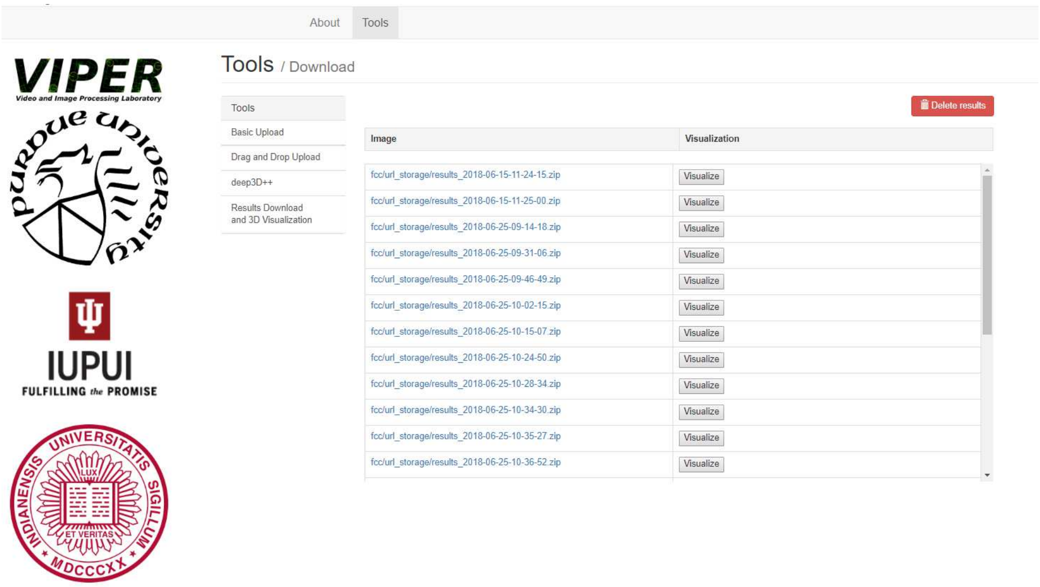
Width: 512 Height: 512

Selected Subvolume Scale:

Width: 177 Height: 201

Fig. 5.6.: Subvolume selecting functionality

A deep learning based nuclei segmentation method, deep 3D+ [53], is implemented in our system. In Figure 5.5, five different segmentation models that trained with different microscopy images are provided. User can process their uploaded data with these models. Also, a image preview window shows the uploaded image. As shown in Figure 5.6, user can also process on a subvolume of the data by specifying a region of interest in the preview window. By pushing the blue ”Process” button, our system will process the data at our computing clusters. Once our system finished the process, the web page will automatically redirected to the result download page. As shown in Figure 5.7, user can download the result or visualize the result immediately in our built in 3D visualization tool. In Figure 5.8, our visualization tool can also provide subvolume visualization and 2D slices visualization.



Tools / Download

Image	Visualization
fcc/url_storage/results_2018-06-15-11-24-15.zip	Visualize
fcc/url_storage/results_2018-06-15-11-25-00.zip	Visualize
fcc/url_storage/results_2018-06-25-09-14-18.zip	Visualize
fcc/url_storage/results_2018-06-25-09-31-06.zip	Visualize
fcc/url_storage/results_2018-06-25-09-46-49.zip	Visualize
fcc/url_storage/results_2018-06-25-10-02-15.zip	Visualize
fcc/url_storage/results_2018-06-25-10-15-07.zip	Visualize
fcc/url_storage/results_2018-06-25-10-24-50.zip	Visualize
fcc/url_storage/results_2018-06-25-10-28-34.zip	Visualize
fcc/url_storage/results_2018-06-25-10-34-30.zip	Visualize
fcc/url_storage/results_2018-06-25-10-35-27.zip	Visualize
fcc/url_storage/results_2018-06-25-10-36-52.zip	Visualize

Fig. 5.7.: Download page of DINAVID

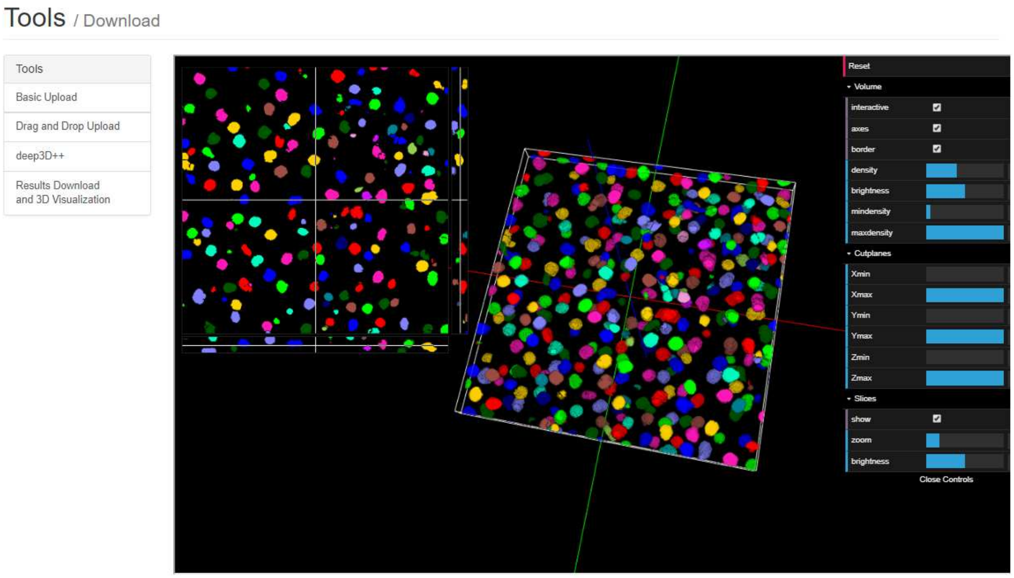


Fig. 5.8.: 3D visualization of DINAVID

6. SUMMARY AND FUTURE WORK

6.1 Summary

In this thesis, we focused on the image analysis on microscopy images including image registration, image synthesis and image segmentation. A 4D image registration that uses combination of rigid and non-rigid registration was described. A Quasi 2D nuclei segmentation was developed convolutional neural networks. We investigated in nuclei image synthesis to solve lack of training data problem. A nuclei image synthesis technique spatially constrained cycle-consistent adversarial network was proposed to generate nuclei image. A 3D segmentation using a combination of binary cross entropy loss and dice loss was presented later. Finally, a multi-task U-Net was described to segmentation nuclei as individual instance. The main contribution of this thesis are as follows:

- 4D Image Registration

  We extended previous work of 3D image registration method to a 4D registration method. The 4D registration method enables fixing motion artifacts in depth of the live tissue and motion artifacts in time dimension. Three dimensional spherical histograms of motion vectors were used to validate our method.

- Image Synthesis

  We proposed a spatial constrained cycle-consistent adversarial network for nuclei image synthesis. This method generates realistic nuclei images with corresponding segmentation labels. This method requires no segmentation labels for training. This method enabled the training of machine learning based techniques for nuclei segmentation.

- 2D Nuclei Segmentation

  We described a 2D CNN segmentation method to segmentation only nuclei from the 3D image volumes that also contains different non-nuclei biological structures. We are able to accurately segment nuclei from 3D image volumes by using our system. Watershed based nuclei counting was able to separate overlapped nuclei and count them.

- 3D Nuclei Segmentation

  We described a 3D CNN segmentation method to segmentation 3D nuclei from the 3D image volumes. A combination of dice loss and binary cross entropy loss were used to train a modifed U-Net. With our SpCycleGAN nuclei data generation, we were able to training our 3D U-Net in a large scale. A Quasi-3D watershed was applied on the segmentation to separate overlapping nuclei. This method achieves promising results in terms of object-based evaluation and voxel-based evaluation

- 3D Nuclei Segmentation

  We also proposed a instance segmentation method, multi-task U-Net. This method generates segmentation mask with corresponding nuclei location map. Using marker-controlled watershed, our method was able to separate overlapping nuclei and minimize over-segmentation of watershed-based technique.

- Distributed and Networked Analysis of Volumetric Image Data (DINAVID)

  We create a Distributed and Networked Analysis of Volumetric Image Data (DINAVID) system. DINAVID is web-based microscopy image analysis system. This system enables biologists to do fast and accurate analysis on microscopy images. After analysis, a 3D visualization of the results can also be viewed in our system.

6.2 Future Work

- Image Registration

  Currently, our registration method is limited to 4D rigid registration due to the need of preserving the original motion of cells in our dataset. In many other microscopy image registration problems, a 4D non-rigid registration method would be used to generate the best results. In the future, we plan to generalize our method to a 4D non-rigid registration technique that can cancel the non-rigid motion artifacts in temporal 3D images.

- Image Synthesis

  Our image SpCycleGAN is able to generate nuclei images without using any manually labeled data. The generated 2D images can be stacked to form 3D volume. Although the characteristic of nuclei is realistic in 2D, shape of the structures are not perfectly defined in 3D. In the future, we would like to expanding our current method to a 3D technique. Also, our current method can be used on other applications such as image de-noising and image restoration.

- Nuclei Segmentation

  Although our nuclei segmentation achieves high accuracy in terms of object-based and voxel-based evaluation, the generalization of our model remains a problem. The characteristic of biological structures varied from different organs and different data acquisitions. A generalized model is hard to obtain due to lack of labeled data. Since our SpCycleGAN can be used to cheaply generate training data for our segmentation training, our current approach is to generate segmentation models for different groups of microscopy images. In the future, we would like to explore more on how to generalize our techniques.

- Distributed and Networked Analysis of Volumetric Image Data (DINAVID)

  We will continue to develop our web-based image analysis system with more features based on the feedback of biologist.

6.3 Publication Resulting From This Work

Journal Papers

1. **C. Fu**, S. Han, S. Lee, D. J. Ho, P. Salama, K. W. Dunn and E. J. Delp, "Three Dimensional Nuclei Synthesis and Instance Segmentation", *To be Submitted, IEEE Transactions on Medical Imaging.*

2. D. J. Ho, **C. Fu**, D. M. Montserrat, P. Salama and K. W. Dunn and E. J. Delp, "Sphere Estimation Network: Three Dimensional Nuclei Detection of Fluorescence Microscopy Images", *To be Submitted, IEEE Transactions on Medical Imaging.*

Conference Papers

1. **C. Fu**, N. Gadgil, K. K Tahboub, P. Salama, K. W. Dunn and E. J. Delp, "Four Dimensional Image Registration For Intravital Microscopy", *Proceedings of the Computer Vision for Microscopy Image Analysis workshop at Computer Vision and Pattern Recognition*, July 2016, Las Vegas, NV.

2. **C. Fu**, D. J. Ho, S. Han, P. Salama, K. W. Dunn, E. J. Delp, "Nuclei segmentation of fluorescence microscopy images using convolutional neural networks", *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 704-708, April 2017, Melbourne, Australia. DOI: 10.1109/ISBI.2017.7950617

3. **C. Fu**, S. Han, D. J. Ho, P. Salama, K. W. Dunn and E. J. Delp, "Three dimensional fluorescence microscopy image synthesis and segmentation", *Proceedings of the Computer Vision for Microscopy Image Analysis workshop at Computer Vision and Pattern Recognition*, June 2018, Salt Lake City, UT.

4. D. J. Ho, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Segmentation of Fluorescence Microscopy Images Using Three Dimensional Convolutional Neural Networks," *Proceedings of the Computer Vision for Microscopy*

Image Analysis (CVMI) workshop at Computer Vision and Pattern Recognition (CVPR), July 2017, Honolulu, HI. DOI: 10.1109/CVPRW.2017.116

5. D. J. Ho, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Detection and Segmentation of Fluorescence Microscopy Images Using Three Dimensional Convolutional Neural Networks", *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 418-422, April 2018, Washington, DC. DOI: 10.1109/ISBI.2018.8363606

6. S. Lee, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Tubule Segmentation of Fluorescence Microscopy Images Based on Convolutional Neural Networks with Inhomogeneity Correction," *Proceedings of the IS&T Conference on Computational Imaging XVI*, February 2018, Burlingame, CA.

7. D. J. Ho, S. Han, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Center-Extraction-Based Three Dimensional Nuclei Instance Segmentation of Fluorescence Microscopy Images," *Submitted To, Proceedings of the IEEE International Symposium on Biomedical Imaging*, April 2019, Venice, Italy.

8. S. Han, S. Lee, **C. Fu**, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei Counting in Microscopy Images with Three Dimensional Generative Adversarial Networks, *To, Appear, Proceedings of the SPIE Conference on Medical Imaging*, February 2019, San Diego, California.

REFERENCES

REFERENCES

[1] D. B. Murphy and M. W. Davidson, *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley-Blackwell, 2012.

[2] J. W. Lichtman and J.-A. Conchello, "Fluorescence microscopy," *Nature Methods*, vol. 2, no. 12, pp. 910–919, December 2005.

[3] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, and D. C. Prasher, "Green fluorescent protein as a marker for gene expression," *Science*, vol. 263, no. 5148, pp. 802–805, February 1994.

[4] T. Stearns, "Green flourescent protein: The green revolution," *Current Biology*, vol. 5, no. 3, pp. 262–264, March 1995.

[5] O. Shimomura, "The discovery of aequorin and green fluorescent protein," *Journal of Microscopy*, vol. 217, no. 1, pp. 3–15, January 2005.

[6] M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning*, vol. 10, no. 4, pp. 128–138, 1988.

[7] E. Wang, C. M. Babbey, and K. W. Dunn, "Performance comparison between the high-speed Yokogawa spinning disc confocal system and single-point scanning confocal systems," *Journal of Microscopy*, vol. 218, no. 2, pp. 148–159, May 2005.

[8] W. Denk, J. H. Strickler, and W. W. Webb, "Two-photon laser scanning fluorescence microscopy," *Science*, vol. 248, no. 4951, pp. 73–76, April 1990.

[9] D. W. Piston, "Imaging living cells and tissues by two-photon excitation microscopy," *Trends in Cell Biology*, vol. 9, no. 2, pp. 66–69, February 1999.

[10] K. W. Dunn, R. M. Sandoval, K. J. Kelly, P. C. Dagher, G. A. Tanner, S. J. Atkinson, R. L. Bacallao, and B. A. Molitoris, "Functional studies of the kidney of living animals using multicolor two-photon microscopy," *American Journal of Physiology-Cell Physiology*, vol. 283, no. 3, pp. C905–C916, September 2002.

[11] F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," *Nature Methods*, vol. 2, no. 12, pp. 932–940, December 2005.

[12] K. Svoboda and R. Yasuda, "Principles of two-photon excitation microscopy and its applications to neuroscience," *Neuron*, vol. 50, no. 6, pp. 823–839, 2006.

[13] E. H. Hoover and J. A. Squier, "Advances in multiphoton microscopy technology," *Nature Photonics*, vol. 7, pp. 93–101, February 2013.

[14] J. Peti-Peterdi, I. Toma, A. Sipos, and S. L. Vargas, "Multiphoton imaging of renal regulatory mechanisms," *Physiology*, vol. 24, no. 2, pp. 88–96, April 2009.

[15] C. Sumen, T. R. Mempel, I. B. Mazo, and U. H. von Andrian, “Intravital microscopy: visualizing immunity in context,” *Immunity*, vol. 21, no. 3, pp. 315–329, September 2004.

[16] M. J. Hickey and P. Kubes, “Intravascular immunity: the host–pathogen encounter in blood vessels,” *Nature Reviews Immunology*, vol. 9, no. 5, pp. 364–375, May 2009.

[17] L. Qu, F. Long, and H. Peng, “3-D registration of biological images and models: Registration of microscopic images and its uses in segmentation and annotation,” *IEEE Signal Processing Magazine*, vol. 32, no. 1, pp. 70–77, January 2015.

[18] B. Zitova and J. Flusser, “Image registration methods: A survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, October 2003.

[19] K. S. Arun and K. S. Sarath, “An automatic feature based registration algorithm for medical images,” *International Conference on Advances in Recent Technologies in Communication and Computing*, pp. 174–177, October 2010.

[20] P. Matula, M. Kozubek, and V. Dvorak, “Fast point-based 3-D alignment of live cells,” *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2388–2396, August 2006.

[21] S. Chang, F. Cheng, W. Hsu, and G. Wu, “Fast algorithm for point pattern matching: Invariant to translations rotations and scale changes,” *Pattern Recognition*, vol. 30, no. 2, pp. 311–320, February 1997.

[22] K. Mkrtchyan, A. Chakraborty, and A. Roy-Chowdhury, “Optimal landmark selection for registration of 4D confocal image stacks in arabidopsis,” *To appear, IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2016.

[23] A. Myronenko and X. Song, “Intensity-based image registration by minimizing residual complexity,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 11, pp. 1882–1891, June 2010.

[24] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. hawkes, “A comparison of similarity measures for use in 2-D-3-D medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 586–595, August 1998.

[25] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, April 1997.

[26] K. S. Lorenz, P. Salama, K. W. Dunn, and E. J. Delp, “Digital correction of motion artefacts in microscopy image sequences collected from living animals using rigid and nonrigid registration,” *Journal of Microscopy*, vol. 245, no. 2, pp. 148–160, February 2012.

[27] K. S. Lorenz, “Registration and segmentation based analysis of microscopy images,” Ph.D. dissertation, Purdue University, West Lafayette, IN, August 2012.

[28] P. Thevenaz, U. E. Ruttimann, and M. Unser, “A pyramid approach to subpixel registration based on intensity,” *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, January 1998.

[29] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.

[30] C. A. Wilson and J. A. Theriot, “A correlation-based approach to calculate rotation and translation of moving cells,” *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1939–1951, July 2006.

[31] S. Yang, D. Kohler, K. Teller, T. Cremer, P. L. Baccon, E. Heard, R. Eils, and K. Rohr, “Nonrigid registration of 3-D multichannel microscopy images of cell nuclei,” *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 493–499, April 2008.

[32] I. H. Kim, Y. C. M. Chen, D. L. Spector, R. Eils, and K. Rohr, “Nonrigid registration of 2-D and 3-D dynamic cell nuclei images for improved classification of subcellular particle motion,” *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 1011–1022, September 2010.

[33] T. Du and M. Wasser, “3D image stack reconstruction in live cell microscopy of Drosophila muscles and its validation,” *Cytometry Part A*, vol. 75, no. 4, pp. 329–343, April 2009.

[34] R. G. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, December 1981.

[35] K. S. Lorenz, P. Salama, K. W. Dunn, and E. J. Delp, “Digital correction of motion artifacts in microscopy image sequences collected from living animals using rigid and nonrigid registration,” *Journal of Microscopy*, vol. 245, no. 2, pp. 148–160, February 2012.

[36] K. W. Dunn, K. S. Lorenz, P. Salama, and E. J. Delp, “IMART software for correction of motion artifacts in images collected in intravital microscopy,” *IntraVital*, vol. 3, no. 1, pp. e28 210:1–10, February 2014.

[37] D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, August 1989.

[38] C. G. Broyden, “The convergence of a class of double-rank minimization algorithms,” *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.

[39] R. Fletcher, “A new approach to variable metric algorithms,” *The Computer Journal*, vol. 13, no. 3, pp. 317–322, 1970.

[40] D. Goldfarb, “A family of variable-metric methods derived by variational means,” *Mathematics of Computation*, vol. 24, no. 109, pp. 23–26, January 1970.

[41] D. F. Shanno, “Conditioning of quasi-Newton methods for function minimization,” *Mathematics of Computation*, vol. 24, no. 111, pp. 647–656, July 1970.

[42] B. Schmid, J. Schindelin, A. Cardona, M. Longair, and M. Heisenberg, "A high-level 3D visualization API for Java and ImageJ," *BMC Bioinformatics*, vol. 11, no. 274, May 2010.

[43] D. J. Ho, C. Fu, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei segmentation of fluorescence microscopy images using three dimensional convolutional neural networks," *Proceedings of the Computer Vision for Microscopy Image Analysis workshop at Computer Vision and Pattern Recognition*, pp. 834–842, July 2017, Honolulu, HI.

[44] S. Rajaram, B. Pavie, N. E. F. Hac, S. J. Altschuler, and L. F. Wu, "SimuCell: A flexible framework for creating synthetic microscopy images," *Nature methods*, vol. 9, no. 7, pp. 634–635, June 2012.

[45] D. Svoboda and V. Ulman, "Generation of synthetic image datasets for time-lapse fluorescence microscopy," *International Conference Image Analysis and Recognition*, pp. 473–482, June 2012.

[46] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2672–2680, December 2014, Montreal, Canada.

[47] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434v2*, January 2016.

[48] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875v3*, December 2017.

[49] A. Osokin, A. Chessel, R. E. C. Salas, and F. Vaggi, "Gans for biological image synthesis," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2252–2261, October 2017, venice, Italy.

[50] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5967–5976, July 2017, Honolulu, HI.

[51] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv preprint arXiv:1703.10593*, pp. 1–16, March 2017.

[52] Y. Huo, Z. Xu, S. Bao, A. Assad, R. G. Abramson, and B. A. Landman, "Adversarial synthesis learning enables segmentation without target modality ground truth," *arXiv preprint arXiv:1712.07695*, pp. 1–4, December 2017.

[53] C. Fu, S. Lee, D. J. Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Three dimensional fluorescence microscopy image synthesis and segmentation," *arXiv preprint arXiv:1801.07198*, pp. 1–9, April 2018.

[54] C. Vonesch, F. Aguet, J. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 20–31, May 2006.

[55] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, January 1979.

[56] W. Niblack, *An introduction to digital image processing.* Prentice-Hall, 1986, vol. 34.

[57] J. Sauvola and M. Pietikäinen, “Adaptive document image binarization,” *Pattern recognition*, vol. 33, no. 2, pp. 225–236, 2000.

[58] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, January 1988.

[59] B. Li and S. T. Acton, “Automatic active model initialization via Poisson inverse gradient,” *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1406–1420, August 2008.

[60] B. Li and S. T. Acton, “Active contour external force using vector field convolution for image segmentation,” *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2096–2106, August 2007.

[61] T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, February 2001.

[62] K. S. Lorenz, P. Salama, K. W. Dunn, and E. J. Delp, “Three dimensional segmentation of fluorescence microscopy images using active surfaces,” *Proceedings of the IEEE International Conference on Image Processing*, pp. 1153–1157, September 2013, Melbourne, Australia.

[63] L. Vincent and P. Soille, “Watersheds in digital spaces: an efficient algorithm based on immersion simulations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, June 1991.

[64] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed., Boston, MA, 2001.

[65] X. Yang, H. Li, and X. Zhou, “Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 11, pp. 2405–2414, November 2006.

[66] A. Dufour, V. Shinin, S. Tajbakhsh, N. Guillen-Aghion, J. C. Olivo-Marin, and C. Zimmer, “Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces,” *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1396–1410, September 2005.

[67] O. Dzyubachyk, W. A. van Cappellen, J. Essers, W. J. Niessen, and E. Meijering, “Advanced level-set-based cell tracking in time-lapse fluorescence microscopy,” *IEEE Transactions on Image Processing*, vol. 29, no. 3, pp. 852–867, March 2010.

[68] J. Cardinale, G. Paul, and I. F. Sbalzarini, “Discrete region competition for unknown numbers of connected regions,” *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3531–3545, August 2012.

[69] G. Paul, J. Cardinale, and I. F. Sbalzarini, “Coupling image restoration and segmentation: A generalized linear model/Bregman perspective,” *International Journal of Computer Vision*, vol. 104, no. 1, pp. 69–93, March 2013.

[70] A. Rizk, G. Paul, P. Incardona, M. Bugarski, M. Mansouri, A. Niemann, U. Ziegler, P. Berger, and I. F. Sbalzarini, “Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh,” *Nature Protocols*, vol. 9, no. 3, pp. 586–596, February 2014.

[71] G. Srinivasa, M. C. Fickus, Y. Guo, A. D. Linstedt, and J. Kovacevic, “Active mask segmentation of fluorescence microscope images,” *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1817–1829, August 2009.

[72] S. C. Zhu and A. Yuille, “Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, September 1996.

[73] Y. Shi and W. C. Karl, “A real-time algorithm for the approximation of level-set-based curve evolution,” *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 645–656, May 2008.

[74] J. O. Cardinale, “Unsupervised segmentation and shape posterior estimation under Bayesian image models,” Ph.D. dissertation, Swiss Federal Institute of Technology in Zurich, Zurich, Switzerland, January 2013.

[75] S. Mallat and S.Zhong, “Characterization of signals from multiscale edges,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.

[76] L. Zhang and P. Bao, “Edge detection by scale multiplication in wavelet domain,” *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1771–1784, December 2002.

[77] Z. Zhang, S. Ma, H. Liu, and Y. Gong, “An edge detection approach based on directional wavelet transform,” *Computers and Mathematics with Applications*, vol. 57, no. 8, pp. 1265–1271, April 2009.

[78] S. Arslan, T. Ersahin, R. Cetin-Atalay, and C. Gunduz-Demir, “Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1121–1131, June 2013.

[79] N. Gadgil, P. Salama, K. Dunn, and E. Delp, “Nuclei segmentation of fluorescence microscopy images based on midpoint analysis and marked point process,” *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 37–40, March 2016, Santa Fe, NM.

[80] Y. He, Y. Meng, H. Gong, S. Chen, B. Zhang, W. Ding, Q. Luo, and A. Li, “An automated three-dimensional detection and segmentation method for touching cells by integrating concave points clustering and random walker algorithm,” *PLoS ONE*, vol. 9, no. 8, pp. e104 437 1–15, August 2014.

[81] O. Cicek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432, October 2016, Athens, Greece.

[82] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. Mok, L. Shi, and P.-A. Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, May 2016.

[83] H. Chen, Q. Dou, L. Yu, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for volumetric brain segmentation," *arXiv preprint arXiv:1608.05895*, August 2016.

[84] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, August 2002.

[85] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, November 1998.

[86] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[87] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proceedings of the Neural Information Processing Systems*, pp. 1097–1105, December 2012, Lake Tahoe, NV.

[88] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, June 2015, Boston, MA.

[89] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[90] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556v6*, April 2015.

[91] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," *Proceedings of the Neural Information Processing Systems*, pp. 1–9, December 2012, Lake Tahoe, NV.

[92] B. Dong, L. Shao, M. D. Costa, O. Bandmann, and A. F. Frangi, "Deep learning for automatic cell detection in wide-field microscopy zebrafish images," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 772–776, April 2015, Brooklyn, NY.

[93] M. Kolesnik and A. Fexa, "Multi-dimensional color histograms for segmentation of wounds in images," *Proceedings of the International Conference Image Analysis and Recognition*, pp. 1014–1022, September 2005, Toronto, Canada.

[94] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, pp. 231–241, October 2015, Munich, Germany.

[95] S. E. A. Raza, L. Cheung, D. Epstein, S. Pelengaris, M. Khan, and N. Rajpoot, "MIMO-Net: A multi-input multi-output convolutional neural network for cell segmentation in fluorescence microscopy images," *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 337–340, April 2017, Melbourne, Australia.

[96] J. Yi, P. Wu, D. J. Hoeppner, and D. Metaxas, "Pixel-wise neural cell instance segmentation," pp. 373–377, April 2018.

[97] A. Arbelle and T. R. Raviv, "Microscopy cell segmentation via adversarial neural networks," pp. 645–648, April 2018.

[98] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 550–566, February 2016.

[99] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, May 2016.

[100] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, pp. 246–253, September 2013, Nagoya, Japan.

[101] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proceedings of the IEEE 2016 Fourth International Conference on 3D Vision*, pp. 565–571, October 2016, Stanford, CA.

[102] D. J. Ho, C. Fu, P. Salama, K. W. Dunn, and E. J. Delp, "Nuclei detection and segmentation of fluorescence microscopy images using three dimensional convolutional neural networks," pp. 418–422, April 2018.

[103] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, July 1994.

[104] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, March 2015.

[105] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

[106] J. L. Clendenon, C. L. Phillips, R. M. Sandoval, S. Fang, and K. W. Dunn, "Voxx: a PC-based, near real-time volume rendering system for biological microscopy," *American Journal of Physiology-Cell Physiology*, vol. 282, no. 1, pp. C213–C218, January 2002.

[107] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, pp. 1–15, December 2014.

[108] S. Lee, P. Salama, K. W. Dunn, and E. J. Delp, “Segmentation of fluorescence microscopy images using three dimensional active contours with inhomogeneity correction,” *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 709–713, April 2017, Melbourne, Australia.

[109] 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, P. Golland, and D. M. Sabatini, “CellProfiler: Image analysis software for identifying and quantifying cell phenotypes,” *Genome Biology*, vol. 7, no. 10, pp. R100–1–11, October 2006.

[110] C. Fu, D. J. Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, “Nuclei segmentation of fluorescence microscopy images using convolutional neural networks,” *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pp. 704–708, April 2017, Melbourne, Australia.

[111] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, July 2006.

[112] D. M. W. Powers, “Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, December 2011.

[113] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, June 2006.

[114] K. U. Barthel, “Volume viewer,” https://imagej.nih.gov/ij/plugins/volume-viewer.html.

VITA

VITA

Chichen Fu was born in Nanchang, Jiangxin Province, China. He received the Bachelor of Science in Electrical Engineering from Purdue University, West Lafayette, Indiana in 2014.

Chichen Fu then joined the Ph.D program at the School of Electrical and Computer Engineering at Purdue University in August 2014. He worked as a research assistant at the Video and Image Processing Laboratory (VIPER) under supervision of Professor Edward J. Delp. Chichen Fu's research interests include image processing, computer vision and deep learning.

He is a student member of the IEEE, the IEEE Signal Processing Society.