Quantification of Multimodal Fluorescence Imaging for Cancer Detection, Guiding Biopsy and Guiding Surgery

Yang Jiang

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:
Eric J. Seibel, Chair

Matthew O'Donnell

Lawrence D. True

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Quantification of Multimodal Fluorescence Imaging for Cancer Detection, Guiding Biopsy and Guiding Surgery

Yang Jiang

Chair of the Supervisory Committee:
Professor Eric J. Seibel
Mechanical Engineering

Fluorescence imaging is emerging for guiding biopsy, early cancer detection and guiding surgery, with the development of new molecularly selective fluorescence agents and fluorescence imaging systems. Accurate quantification of fluorescence imaging is needed to provide diagnostic information, such as tumor stage, tumor location etc. Several fluorescence quantification methods have been addressed in this dissertation, to correlate fluorescence with histology, which is the gold-standard for cancer detection. First, an automatic ratiometric algorithm, target to background (T/B) ratio is developed for esophageal cancer classification and guiding biopsy using single fluorescence labeled peptide. This method may be generally used for other cancer detection in GI tract. Second, a registration pipeline is established to correlate near infrared fluorescence images to histological images, which may contribute to evaluating the sensitivity and specificity of new fluorescence labeled contrast agents in cancer detection. Then T/B ratio is applied to quantify multimodal fluorescence imaging with multiple fluorescence labeled peptides for guiding biopsy and early neoplasia detection in esophagus. Lastly, advanced deep learning algorithms are applied to achieve real-time, automatic frame selection and T/B ratio calculation, which allows computer-aided quantitative fluorescence endoscopy for cancer surveillance. The methodology developed in this work has been used for quantifying multimodal fluorescence imaging, which can assist clinicians in making decisions.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Here I would like to express my sincere gratitude towards people who have helped me during my doctoral journey. I could not have done it without any of them.

First of all, I would like to thank my advisor, Dr. Eric J. Seibel, for being the best mentor I could have asked for. His persistence, passion and devotion to his work is unparalleled and inspiring. I appreciate the training of independent and critical thinking he has gifted me. Outside research, I would hope to inherit his optimism and love for simple joys in life.

The support and help from my supervisory committee are also indispensable during my doctoral study. I am lucky to have another three amazing supervisory committee members, Dr. Lawrence D. True, Dr. Matthew O'Donnell and Dr. Daniel T. Chiu. I'm grateful for the support and guidance they have given me over the years.

I have also received so much help and support from the lab staff, peers and collaborators. My research work would not be possible without Cathy's talented fabrication of SFE and her technical support; Matt has taught me all the hardware fundamentals that supported my research. Len and Mark have given me so much technical guidance as well. I am thankful for the legacy of the great work done by our alumni Yuanzheng. Saniel, Yaxuan, Chen and Manuja have been the support buddies through trouble-shooting and graduate school struggles. My research also would not succeed without the brilliant collaborators from Dr. Tom Wang's group at University of Michigan, Ann Arbor and Dr. James Olson Lab at Fred Hutchinson Cancer Research Center.

Last but not least, I would like to thank my family for supporting me through the whole journey, emotionally and financially. My wife, Mengyuan and our loving dogs, Izzie and Mere have been with me every step of the way and have been my source of joy in difficult times. I look forward to meeting another fruit of this journey, our first child, who is kicking in the womb as I write this.

This work was funded by NIH R01 CA200007, U54 CA163059, NCI SBIR R43 CA211086 and NIH R01 EB016457.

DEDICATION

to my first child, Ari

Chapter 1. INTRODUCTION AND BACKGROUND

1.1 CLINICAL PROBLEMS

1.1.1 *Cancer*

Cancer is the second leading cause of death in the United States. Cancer is a group of diseases that involve uncontrolled growth. It differs from benign tumors in its potential to invade or spread to other parts of the body. Numerous factors are associated with increase in cancer occurrences, including genetic factors and lifestyle factors that can be modified. According to American Cancer Society, it is predicted that more than 1.8 million new cancer cases to be diagnosed in 2020, and about 606,250 to die of cancer in 2020 in US (approximately 1,660 deaths per day) [1]. For example, among gastrointestinal (GI) cancers, esophagus cancer is estimated to have 18,400 new cases with 16,170 estimated deaths; colon cancer with 104,610 new cases with 53,200 deaths; and biliary cancer 11,980 new cases and 4,090 deaths [1].

The 5-year relative survival rate for cancers has increased significantly over the past few decades. This is due to both earlier diagnosis for some cancers as well as advancement in treatment protocol. Cancer death rates have been used as the measurement for progress against disease and we witnessed a substantial and continuous drop since 1991 [1]. As of 2017, the cancer death rate has dropped to 152 per 100,000 people [1]. This is considered the result of reduction in smoking, as well as advancement in early detection and treatment protocol.

1.1.2 *Early cancer detection*

Early detection of cancer with proper treatment will greatly improve the outcome of patients. The stage of the cancer describes the size of the primary tumor and the spread of the cancer at diagnosis. Clinicians mainly use a TNM staging system that categorize cancer into 0, I, II, III, IV stages based on the size or extent of the primary tumor (T) and absence or presence of regional lymph node involvement (N) and distant metastases (M). Recent advancements in early cancer detection have contributed significantly to the improvement in cancer survival rate. Normally speaking, the earlier detection, the better. As shown in Figure 1.1, if prostate cancer, ovarian cancer and breast cancer can be detected at STAGE 1, the 5-year survival rate will improve up to over 90%, compared to less than 30% if detected at Stage 3 (Figure 1.1) [1]. Similarly, for esophagus cancer, cancer survival rate drops 42% if detected at Stage 1 compared to Stage 3. Early detection is essential to improving cancer patient prognosis. Techniques, with increasing sensitivity and specificity, that can detect cancer early and accurately are needed.

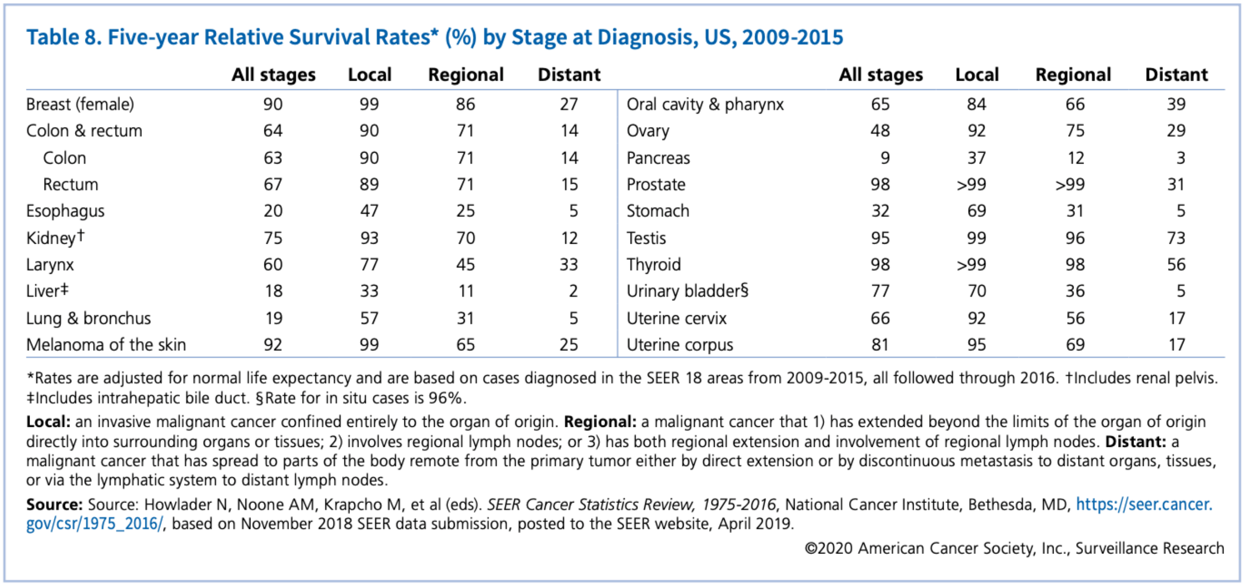


Table 8. Five-year Relative Survival Rates* (%) by Stage at Diagnosis, US, 2009-2015

	All stages	Local	Regional	Distant		All stages	Local	Regional	Distant
Breast (female)	90	99	86	27	Oral cavity & pharynx	65	84	66	39
Colon & rectum	64	90	71	14	Ovary	48	92	75	29
Colon	63	90	71	14	Pancreas	9	37	12	3
Rectum	67	89	71	15	Prostate	98	>99	>99	31
Esophagus	20	47	25	5	Stomach	32	69	31	5
Kidney†	75	93	70	12	Testis	95	99	96	73
Larynx	60	77	45	33	Thyroid	98	>99	98	56
Liver‡	18	33	11	2	Urinary bladder§	77	70	36	5
Lung & bronchus	19	57	31	5	Uterine cervix	66	92	56	17
Melanoma of the skin	92	99	65	25	Uterine corpus	81	95	69	17

*Rates are adjusted for normal life expectancy and are based on cases diagnosed in the SEER 18 areas from 2009-2015, all followed through 2016. †Includes renal pelvis. ‡Includes intrahepatic bile duct. §Rate for in situ cases is 96%.

Local: an invasive malignant cancer confined entirely to the organ of origin. **Regional:** a malignant cancer that 1) has extended beyond the limits of the organ of origin directly into surrounding organs or tissues; 2) involves regional lymph nodes; or 3) has both regional extension and involvement of regional lymph nodes. **Distant:** a malignant cancer that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis to distant organs, tissues, or via the lymphatic system to distant lymph nodes.

Source: Source: Howlader N, Noone AM, Krapcho M, et al (eds). *SEER Cancer Statistics Review, 1975-2016*, National Cancer Institute, Bethesda, MD, https://seer.cancer.gov/csr/1975_2016/, based on November 2018 SEER data submission, posted to the SEER website, April 2019.

©2020 American Cancer Society, Inc., Surveillance Research

Figure 1.1 5-year relative survival rates by stage at diagnosis, US, 2009 – 2015 [1]

1.2 CURRENT CLINICAL CANCER DETECTION

Current clinical cancer detection primarily relies on laboratory test and imaging techniques.

1.2.1 *Laboratory test*

As one of the most standard testing used clinically, blood test can detect blood protein, and cancer-specific biomarkers (for example, prostate-specific antigen testing for prostate cancer) or tumor cells in serum. However, except for blood cancers, blood testing, as well as similar urine testing, suffers from low sensitivity and specificity.

Biopsy is another main way to detect most types of cancer. Biopsy is often conducted with image guiding when the tumor site is deeper inside the body and hard to locate by touching. However, biopsy is considered invasive and thus comes with potential complications from the procedure, including infection, pain and bleeding etc. Moreover, the inter- and intra-tumor heterogeneity poses a critical challenge to the accuracy of biopsy results because biopsies can fail to capture the complete landscape of the tumor.

1.2.2 *Imaging techniques*

Different imaging techniques have been used for cancer detection, including X-ray, mammography, computed tomography (CT), magnetic resonance imaging (MRI), Positron Emission Tomography (PET), ultrasound, and endoscopy. Different image techniques can offer different yet complimentary information for detecting and staging cancer. For example, for esophageal cancer, chest CT can provide important information regarding tumor size, lymph node involvement, and potential metastases. However, it by itself suffers from low sensitivity [2]. MRI can be incorporated for more accurate staging, as well as PET, which has been found particularly helpful when detecting metastatic diseases (stage IV) [3]. Some imaging modalities have been developed to some specific type of cancer. For example, specific to such GI cancers, modalities like endoscopic ultrasound (EUS), an echoendoscope passed directly into esophagus to visualize the histological layers of the esophagus, has been invented and studied [4]. However, additional biopsies are needed to confirm suspected tissues that are identified through these imaging techniques. Meanwhile, the low detection sensitivity of these imaging techniques limits their ability to detect cancer in its early stage [5].

1.3 MOLECULAR IMAGING

Molecular imaging is a promising technique in both cancer detection and treatment by detecting biomarkers specific to cancer progression. Over the past two decades, the developments in molecular imaging, especially fluorescence imaging, enable the visualization of subcellular structures and biological processes with a higher spatial and temporal resolution compared to conventional imaging modalities [6, 7]. New fluorescence imaging agents, coupled with advanced fluorescence devices, promote the field of cancer detection using fluorescence imaging [8]. Moreover, as a real-time detection modality, fluorescence imaging can provide direct visualization of tumor for intraoperative surgery and highlight malignancies that were once hidden from resection. For surgical resection, a complete resection is highly correlated to clinical outcomes. Any residual tumor may result in increased rates of local recurrence and worse clinical outcomes [9, 10]. More and more clinical trials have been investigated using fluorescence imaging for cancer detection and image guided surgery [11].

1.3.1 *Fluorescent imaging agents*

Several fluorescence contrast agents are being developed and used in vivo during clinical studies, which are based on 3 different targeting mechanisms.

1) Passive targeting

   Molecules of certain sizes (typically nanoparticles, and macromolecular drugs) tend to accumulate in tumor tissue much more than they do in normal tissues. This phenomenon is referred as enhanced permeability and retention effects (EPR). A small number of passive targeting contrast agents have been approved by FDA, including methylene blue, sodium fluorescein and indocyanine green (ICG). ICG is the first and only clinically approved fluorophore in near infrared (NIR) range, which have been widely used in angiography, lymphography. Identification of sentinel lymph nodes (SLN) using ICG assists assessment of lymph node metastases [12]. However, the use of passive targeting fluorophores provides non-specific signals and may not be ideal for cancer detection

2) Metabolic targeting

   5-aminolevulinic acid (5-ALA) has been recently approved by FDA as an intraoperative imaging agent for the malignant glioma resection. The high metabolic activity in tumor cells process 5-ALA to fluorescent protoporphyrin IX (PpIX) via the haeme-synthesis pathway. This substrate has an absorption peak of 405 nm, and an emission peak of 635nm, which can be directly visualized during intraoperative surgery. However, 5-ALA can only be used to visualize surface tumor. The poor penetration of light in visible range limited the detection of deep tumor.

3) Molecular targeting

Imaging agents have been developed for targeting cell-surface receptors such as EGFR and HER2. By using targeted fluorophores, the sensitivity and specificity of detection is expected to increase [13]. Researchers have focused on developing new fluorescence contrast agents by conjugating targeting agents to NIR fluorophore, thus developing cancer-selective fluorophores with deep tissue penetration. In this work, we are using a panel of fluorescent labeled peptides that are specific for overexpressed cell surface targets, including EGFR, and ErbB2, for early esophageal neoplasia detection. A chlorotoxin (CTX) conjugated to a derivative of ICG have been evaluated for brain tumor detection for image guided surgery.

1.4 MULTIMODAL SCANNING FIBER ENDOSCOPE (MMSFE)

Advanced fluorescence imaging instruments have been developed. Our lab is developing a multimodal scanning fiber endoscope. The mmSFE systems are specifically designed for wide field of view, high resolution, and real-time multiple fluorescence imaging, which can be used for cancer detection, guiding biopsy and surgery. Three diode lasers :488, 642, 785nm are combined and sent to a single mode optical fiber, which is driven by a piezo tube actuator and scans in a spiral pattern (Figure 1.2 (a)). Fluorescence and reflected light are then collected by a concentric ring of high numerical aperture (NA) multimodal optical fibers (Figure 1.2 (b)). The mmSFE has an overall flexible shaft diameter of 1.2 to 2.4 mm and a rigid tip length of 9 mm (Figure 1.3(c)). The collected light is separated into 4 channels, with 3 fluorescence channels and 1 reflectance channel. The system is optimized for 3 fluorophores, FITC, Cy5 and IRDye800, which have minimal emission spectral overlap (Figure 1.3).

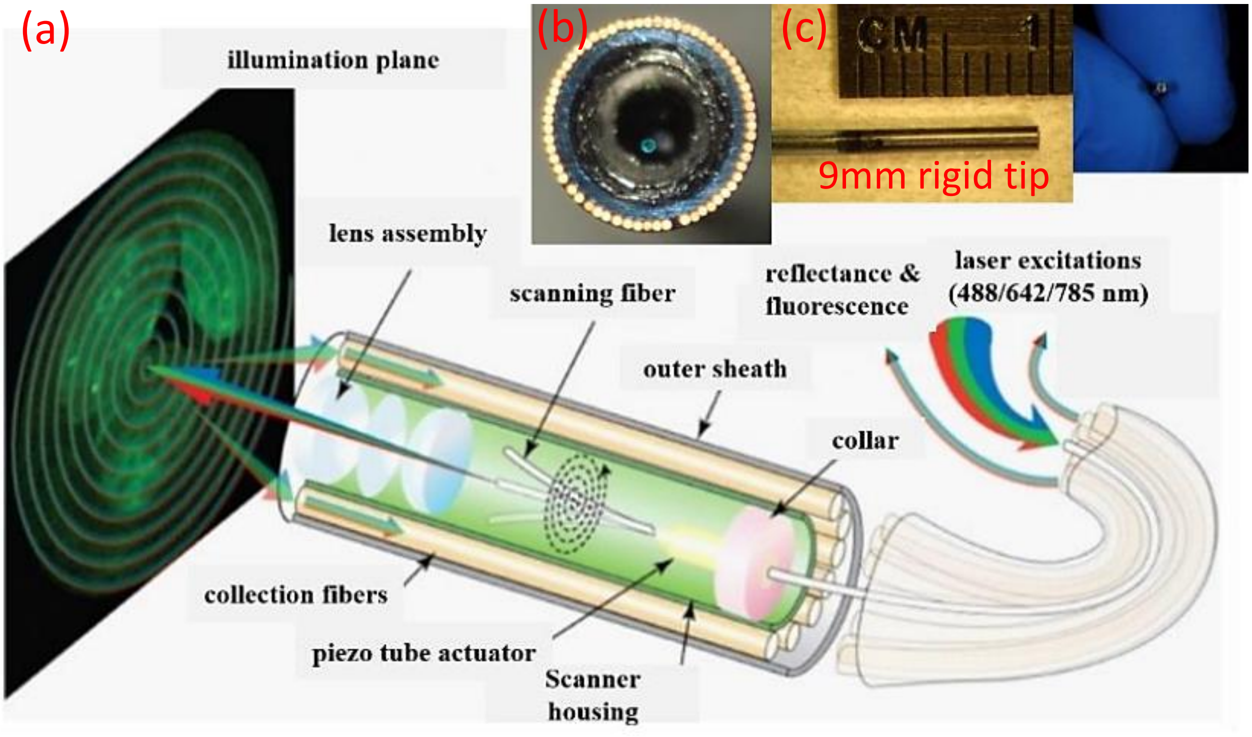
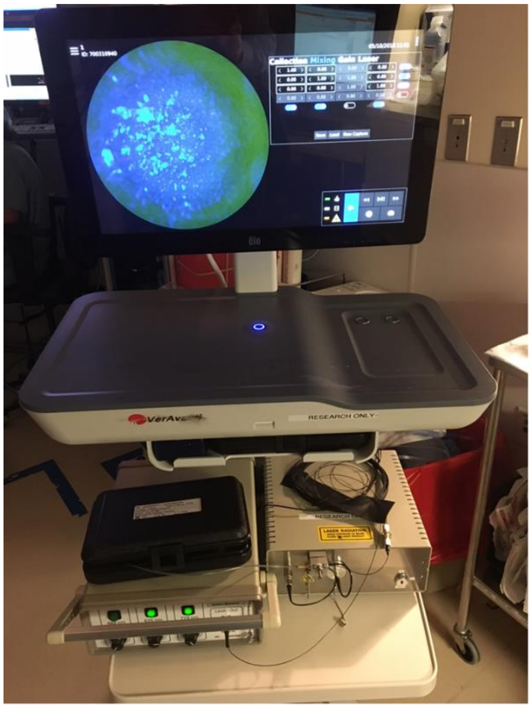


Figure 1.2 Schematic illustration of mmSFE



Channel	Description	Excitation	Emission
1	Fluorescence - FITC	489nm	525nm
2	Reflectance	638nm	
3	Fluorescence - IRDye800	774nm	789nm
4	Fluorescence - Cy5	649nm	670nm

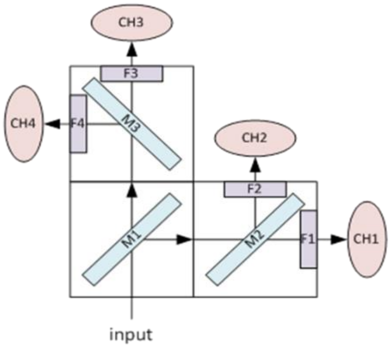


Figure 1.3 mmSFE: 3 fluorescence channel + 1 reflectance channel, optimized for FITC, Cy5 and IRDye800

Sensitivity of the mmSFE is defined as the lowest concentration of fluorescent dye molecules that can be detected. The sensitivity for mmSFE was tested using microhematocrit capillary tubes (1mm ID). Fluorescence dye ICG, Cy5 and FITC was dissolved and diluted in Dimethyl sulfoxide (DMSO). Dilutions covered the range from 100 micromolar (μmol/L) to 1 nanomolar (nmol/L) for all dyes. All measurements were performed by imaging the dyes in capillary tubes, with mmSFE probe 10mm away perpendicular to tube surface (Figure 1.4). As can be seen, the sensitivity is 1nM for ICG, 10nM for Cy5 and FITC.

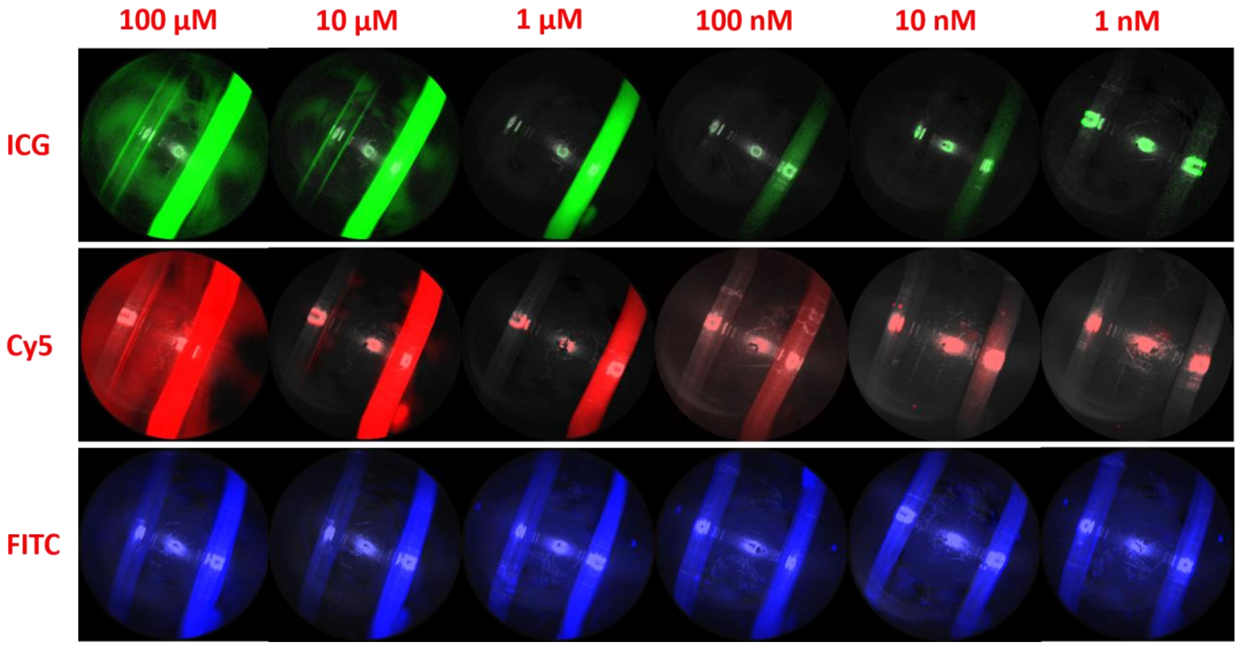


Figure 1.4 mmSFE sensitivity detection, 1nM for ICG, 10nM for Cy5 and FITC

1.5 NEED FOR FLUORESCENCE QUANTIFICATION

In parallel with the development of fluorescence contrast agents and instrumentations, accurate quantification is needed for fluorescence imaging. There are no standards for clinical use of fluorescence. Different fluorescence imaging systems are currently implemented using a variety of cameras, illumination sources, data processing methods and other experimental parameters [14]. To provide quantitative fluorescence imaging,

(1) System calibration is needed before performing clinical studies.

(2) Variable parameters, such as variations in camera distance and angle from tissue, variations in optical tissue properties and ambient light, need to be addressed before analyzing the clinical data [15, 16].

(3) Advanced methods are needed to correlate fluorescence with histology, class-wise and pixel-wise.

1.6 GOAL

The goal of this research is to quantitatively analyze multimodal fluorescence imaging for cancer detection, image guided biopsy and image guided surgery. Specifically, the analysis is conducted on clinical study using mmSFE with fluorescence labeled peptides for early cancer detection and guided biopsy in esophagus. A ratiometric method, target to background (T/B) ratio, is applied to quantify fluorescence signals (Chapter 2, 4). This T/B ratio is calculated from the average intensities of the segmented fluorescence targets and the neighboring background regions, which provides a relative quantification of the differentially expressed biomarkers in esophageal cancer. To achieve an automatic and real-time T/B calculation, deep learning based algorithm are proposed (Chapter 5). In addition, a registration framework is established to correlate NIR fluorescence imaging to histological images, thus evaluating fluorescence agents in delineating tumor core and margins (Chapter 3). In research, this work will develop image-based diagnostic protocols for assessing the sensitivity and specificity of the multimodal fluorescence imaging in cancer detection and fluorescence guided surgery. In clinical settings, this work will provide a computer aided diagnostic (CAD) system for guiding biopsy and help surgeons make decisions.

This dissertation can be divided into 4 parts as follows:

- Single fluorescence imaging analysis for guiding biopsy and esophageal neoplasia detection (Chapter 2)
- Pixel-wise comparison between NIR fluorescence and histological images (Chapter 3)
- Multimodal fluorescence imaging (multiple fluorescence + reflectance) analysis for early cancer detection and guiding biopsy in esophagus (Chapter 4)
- Two-step deep learning models for automatic suspicious frame selection and real-time T/B calculation (Chapter 5).

Chapter 2. TOWARD REAL-TIME QUANTIFICATION OF FLUORESCENCE MOLECULAR PROBES USING TARGET/BACKGROUND RATIO FOR GUIDING BIOPSY AND ENDOSCOPIC THERAPY OF ESOPHAGEAL NEOPLASIA

Multimodal endoscopy using fluorescence molecular probes is a promising method of surveying the entire esophagus to detect cancer progression. In this chapter, we developed a Chan-Vese based algorithm to segment fluorescence targets, and subsequent morphological operations to generate background, thus calculating T/B ratios, potentially to provide real-time guidance for biopsy and endoscopic therapy. Furthermore, an automatic computer-aided diagnosis (CAD) algorithm can be applied to the recorded endoscopic video and the overall T/B ratio is calculated for each patient.

2.1 BACKGROUND

Esophageal adenocarcinoma (EAC) possesses a 5-year survival rate less than 18% in US with increasing incidence [17, 18]. Barrett's esophagus (BE) is regarded as a premalignant condition of EAC. BE is defined as the metaplastic change of esophagus, in which normal squamous epithelium has been transformed to columnar epithelium containing goblet cells [19]. The transformations from BE to EAC progress through low-grade dysplasia (LGD) and high-grade dysplasia (HGD) [20]. Early detection and treatment of EAC at its premalignant stages could increase the survival rate of the patients. Clinically, endoscopic surveillance is often recommended to BE patients. Current standard protocol is applying high resolution white light endoscopy with four-quadrant biopsies at 1-2 cm intervals [21]. However, this approach is intensive and expensive. Due to the lack of visualization of surface abnormality present in dysplastic lesions, extensive biopsies are randomly taken. The histopathological processing of these biopsies provides slow feedback to the endoscopist, often 1 to 2 weeks. Since only a small fraction of the surface area is sampled (4% to 6%), it is still an inaccurate evaluation of dysplasia or EAC due to the sampling error [22]. Therefore, advanced imaging for more comprehensive biopsy guidance and more rapid and accurate detection of dysplastic tissue progressing to EAC is highly demanded.

Several advanced endoscopic imaging techniques have been investigated to detect dysplasia associated with BE [23, 24, 25]. Autofluorescence imaging (AFI) is the phenomenon that endogenous substances, such as collagen and flavins, emit longer wavelength light after being excited with short wavelength light. By distinguishing among different autofluorescence patterns highlighted by AFI, an experienced endoscopist can find more dysplastic lesions from the background of normal squamous epithelium. Boerwinkel et al. showed that AFI possesses an 89% sensitivity of HGD and EAC yet suffers from high false-positive rates of 86% [26]. This is inconsistent to their previous findings concluding that earlier commercial version of AFI had limited ability in detecting dysplasia lesions [23]. Narrow band imaging (NBI) illuminates the mucosa in blue band using laser light around 415 nm, which targets hemoglobin absorption, to enhance the visualization of vascular structures at the mucosal surface [27]. Wolfsen et al. showed NBI can identify patients with dysplasia with significantly (50%) fewer biopsy samples [24]. Recent studies showed that detailed inspection with NBI after AFI reduced the false positive rate from 81% to 26% [28]. Overall, both NBI and AFI use non-specific contrast mechanisms to detect dysplasia lesions, which suffer from low specificity. Angle-resolved low coherence interferometry (a/LCI), a type of optical biopsy, measures the nuclear morphology changes to evaluate dysplastic changes in the esophageal mucosa. However, this technique relies on point sampling, which is not a method of interrogating the entire mucosa having BE [29]. Volumetric laser endomicroscopy (VLE) and tethered capsule endomicroscopy are comprehensive imaging techniques, which can provide frequency domain optical coherence tomography (OCT) imaging from a minimally invasive side-viewing scope [25,30]. However, they can only provide cross-sectional structural data, without molecular or chemical contrast which may be required for detecting cells progressing from LGD to HGD.

Molecular imaging is an emerging technique that can provide molecular biology information specific to cancer progression, which is showing promise in the management of Gastrointestinal (GI) cancers [31, 32]. Optical imaging techniques, such as wide-field endoscopy or smaller field of view endomicroscopy, are used to measure relative intensity of fluorescently labeled probes that are associated with EAC. Using a fluorescence agent that indicates high metabolic activity of neoplastic EAC ex vivo, Leggett et al. achieved a sensitivity of 79% and a specificity of 77% using probe-confocal laser endomicroscopy (pCLE) [33]. However, the fluorescent marker they applied, 2-NBDG, is a non-targeted fluorescence agent and limited by non-specific binding. Bird-Lieberman et al. used fluorescence endoscopy with fluorescently labeled lectin to detect the changes of glycan following the progression from BE to EAC in situ [34]. However, this study was limited by the small number of patient tissue samples and failed to report the sensitivity and specificity of the proposed technique in targeting dysplastic lesions compared to the state of the art. Moreover, lectin is a negative contrast agent, which is limited by false positives due to low specific binding in vivo because of overlying mucus and mucosal folds. By discovering a fluorescent peptide that specifically binds to EAC neoplasia, Sturm et al. attained 75% sensitivity and 97% specificity with first-in-human study using pCLE imaging and histological correlation [35].

By using the same molecularly-specific fluorescence peptide as Sturm et al. [32] and including more subjects (50 subjects) in the study, Joshi et al. [35] achieved 76% sensitivity and 94% after post-processing the fluorescence video frames from a wide-field custom fluorescence endoscope. Joshi et al. assessed the feasibility of using fluorescence labeled target intensity compared to the background intensity to localize HGD and EAC with histological correlation for computer-aided diagnosis. However, a major limitation was the slow image processing that was done after the endoscopic procedure to calculate diagnostic T/B ratios. This computational limitation would not allow the endoscopist to use T/B analysis as an indicator during endoscopic procedure to guide biopsy or resection. Real-time quantification and mapping of the fluorescence image is required as a navigation guide for sampling the highest-risk regions for dysplasia, EAC, as well as providing direct feedback to surgeons when removing these areas and neoplasia by endoscopic mucosal resection (EMR). Moreover, in future applications of surveillance in rural clinics, automatic interpretation of fluorescence images is strongly desired to assist in making diagnostic decisions. Qi et al. achieved automatic interpretation of endoscopic OCT images to detect dysplasia in BE with a sensitivity of 82%, a specificity of 74% [36]. However, to the best of our knowledge, there is no image processing algorithm for wide-field fluorescence endoscopy to (1) guide the endoscopist to the most suspicious regions in BE during interventions of biopsy and EMR in real-time, and (2) to automatically analyze multimodal videos to classify these regions as high-risk dysplastic or neoplastic lesions.

One of the main challenges to achieve (1) real-time guidance and (2) automatic CAD of BE patients using fluorescence endoscopy is to select an accurate and fast algorithmic approach to label the foreground targets in each frame of the fluorescence video. Due to the low contrast "patchy" appearance of the fluorescence targets and non-uniform intensity in both targets and background, the segmentation algorithms using a single threshold value, such as Ostu's algorithm, show unsatisfactory results. These limitations were demonstrated by Chan et al. when applying to fluorescence images [37]. Otsu's algorithm assumes a bimodal intensity histogram of images, while the intensity histograms of fluorescent images are normally unimodal due to large fluorescent signal variation. When applying a single threshold to these unimodal-histogram fluorescent images, small targets with lower intensity will be mistakenly excluded from the labeled target region. On the other hand, the Chan-Vese algorithm [38] can adapt the boundary of the targets iteratively and locally, and thus this algorithm is more robust in noisy low-contrast images. Moreover, efficient processing of the Chan-Vese algorithm allows for real-time display of measured and highlighted regions. To our knowledge, this algorithm has never been applied on fluorescence endoscopic video.

In this chapter, we describe an Chan-Vese based approach to achieve highly efficient T/B calculation from florescence videos and demonstrate a method of highlighting regions of interest (ROIs) that score above our T/B threshold for direct feedback to the endoscope user (part 1). In addition, we develop an automatic CAD algorithm for classifying EAC/HGD from recorded fluorescence videos of esophageal endoscopy with molecular probes (part 2). In recorded videos from 50 subjects, we show that a surveillance protocol based on this automated algorithm using fluorescently-labelled peptides specific for BE neoplasia is capable to classify pre-cancerous and neoplastic (HGD and EAC) regions with 92% sensitivity and 75% specificity using histological confirmation.

2.2 MATERIALS AND METHOD

2.2.1 *Patients and endoscopic procedure*

The primary data for the analyses were described by Joshi et al. [35]. Fifty patients with history of BE participated in the diagnosis of HGD/EAC and its endoscopic treatment at the University of Michigan. Fluorescence and reflectance videos were collected using a customized Olympus endoscope (GIF-Y0029) having 20 frames per second and 140° maximum light collection angle to the camera. Peptide ASY-NYDA, specific binding to early Barrett’s neoplasia, is labeled with fluorescein isothiocyanate (FITC), hereafter called ASY*-FITC [29]. Topical (sprayed) ASY*-FITC peptide was administrated. After short incubation time (5 minutes) and rinsing to remove unbounded peptides, a time-series of white-light, fluorescence, and reflectance videos were collected on selected segments of BE. After clinical evaluation, biopsies and EMR tissue specimens were removed from the same BE segments. The tissue was sectioned in 2-mm intervals and a diagnosis was determined from two gastrointestinal pathologists. Disease classification was recorded as positive (HGD or EAC) or negative (squamous, BE, gastroesophageal junction or LGD) based on histological interpretation, see Table 2 within Joshi et al., 2016 [35].

2.2.2 *Part 1: T/B ratio calculation*

T/B ratio was calculated for each frame in the recorded wide-field fluorescence videos. All these videos were recorded in high resolution (1920×1080) with 8-bit intensity using an Olympus image digitizer. The overall procedure to calculate T/B ratio is illustrated in Figure 2.1.

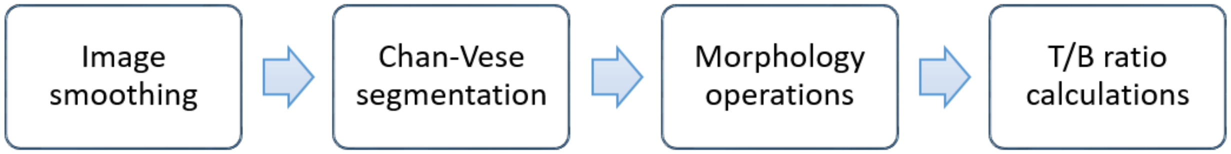


Figure 2.1 T/B ratio calculation procedure for fluorescence image.

For each frame, an image-smoothing step is first performed using a Gaussian filter with 5*5 window. After smoothing, ROIs were automatically segmented using Chan-Vese algorithm. After segmentation, a series of morphology operations were applied to remove noise pixels and to acquire the background pixels surrounding the targets. Then the T/B ratio was calculated as the ratio between the average intensity in the targets and the average intensity in the background. We describe the two main steps in detail below.

2.2.2.1 Segmentation using Chan Vese algorithm [38]

The Chan-Vese algorithm uses an active contour model, which initializes the segmentation with a contour and then evolves this contour via a level set method. A level set function defines the edge of contour. For example, for level set function ϕ (i,j,t), where (i, j) is the pixel coordinate, t is the evolution time step, the edge contour is defined to be the zero level set s.t. ϕ (i,j,t)= 0, while the regions inside and outside contour are given by $\phi > 0$ and $\phi < 0$. The goal of this algorithm is to evolve the level set function to minimize the energy function (Equation 2.1)

$$
\begin{aligned}
\mathrm{F}(\phi) &= \mu\int_{\Omega} |\nabla H(\phi)|\, dx + v\int_{\Omega} H(\phi)dx \\
&+ \lambda_1 \int_{\Omega} |I - c_1|^2\, H(\phi)dx + \lambda_2 \int_{\Omega} |I - c_2|^2\, (1 - H(\phi))dx
\end{aligned}
$$

Equation 2-1

where H is the Heavyside function, ϕ is the level set function, *I* is the image needed to be segmented and Ω is the image domain. c_1 and c_2 are the average image intensity in the regions where $\phi > 0$ and $\phi < 0$ respectively (Equation 2.2, 2.3). μ, ν, λ_1, λ_2 are the regulation parameters for the energy function. μ penalizes the length of the edge contour, while ν is the penalty on the total area inside the contour. λ_1 and λ_2 controls the variance of pixel intensity inside and outside the contour.

$$
c_1 = \frac{\int_{\Omega} I(x)H(\phi)dx}{\int_{\Omega} H(\phi)dx}
$$

Equation 2-2

$$
c_2 = \frac{\int_{\Omega} I(x)(1 - H(\phi))dx}{\int_{\Omega} (1 - H(\phi))dx}
$$

Equation 2-3

To minimize the energy function and to evolve the level set boundary, a semi-implicit gradient descent method is applied [35]. The derived numerical and discretized gradient for the level set function (Equation 2.4) minimizes the energy function iteratively. The level set function is first initialized with a circular function. At each iteration, ϕ_{n+1} is calculated based on Equation 2.4, while c_1 and c_2 are calculated from Equation 2.2, 2.3. The iteration will be terminated when the relative change between ϕ_{n+1} and ϕ_n is smaller than a pre-set threshold or the maximum iteration number is reached, whichever is achieved earlier.

$$\begin{aligned}\frac{\phi_{i,j}^{n+1}-\phi_{i,j}^{n}}{\Delta t}=\delta_{\mathrm{h}}(\phi_{i,j}^{n})\frac{\mu}{h^{2}}&\left[\Delta_{-}^{x}\left(\frac{\Delta_{+}^{x}\phi_{i,j}^{n+1}}{\sqrt{(\Delta_{+}^{x}\phi_{i,j}^{n})^{2}/h^{2}+(\phi_{i,j+1}^{n}-\phi_{i,j-1}^{n})^{2}/(2h)^{2}}}\right)\right.\\&\left.+\Delta_{-}^{y}\left(\frac{\Delta_{+}^{y}\phi_{i,j}^{n+1}}{\sqrt{(\phi_{i+1,j}^{n}-\phi_{i-1,j}^{n})^{2}/(2h)^{2}+(\Delta_{+}^{y}\phi_{i,j}^{n})^{2}/h^{2}}}\right)\right]\\&-\delta_{\mathrm{h}}(\phi_{i,j}^{n})\Big(\nu+\lambda_{1}\big(I_{i,j}-c_{1}(\phi^{n})\big)^{2}-\lambda_{2}\big(I_{i,j}-c_{2}(\phi^{n})\big)^{2}\Big)\end{aligned}$$

Equation 2-4

Where δ_h is the smoothed delta function (Equation 2.5), Δ_- and Δ_+ are the spatial finite difference (Equation 2.6)

$$\delta_{\mathrm{h}}(x)=\frac{1}{\pi}\frac{h}{h^{2}+x^{2}}$$

Equation 2-5

$$\begin{aligned}\Delta_{+}^{x}\phi_{i,j}^{n}&=\phi_{i+1,j}^{n}-\phi_{i,j}^{n} & \Delta_{-}^{x}\phi_{i,j}^{n}&=\phi_{i,j}^{n}-\phi_{i-1,j}^{n}\\\Delta_{+}^{y}\phi_{i,j}^{n}&=\phi_{i,j+1}^{n}-\phi_{i,j}^{n} & \Delta_{-}^{y}\phi_{i,j}^{n}&=\phi_{i,j}^{n}-\phi_{i,j-1}^{n}\end{aligned}$$

Equation 2-6

2.2.2.2 Morphology and T/B ratio calculation

The resulting level set function ϕ defines the regions inside and outside the edge contour, which can be transformed into a binary image. Erosion was then applied to smooth the binary image and to remove isolated noise pixels. The default setting is to remove noise smaller than 10 pixels. Then dilation was carried out to generate a 30 pixel-wide background band surrounding the targets. Average intensities of the segmented targets and the neighboring background bands are used for T/B ratio calculation to be able to match results from Joshi et al. [35].

Using implementation of the procedure above, we measured the running time of T/B ratio calculation for each frame to evaluate its ability to be used for real-time biopsy guidance.

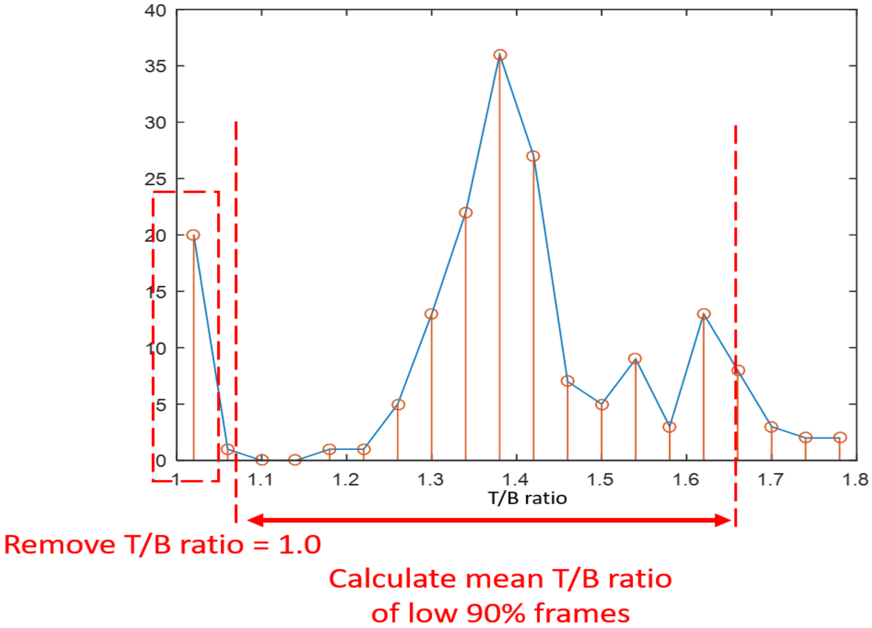
2.2.3 *Part 2: Automated CAD for HGD and EAC*

After T/B ratio was calculated for each frame in the fluorescence videos, a histogram-based method was adopted to remove artificial frames (described below) and to determine the T/B ratio of the patients. Then a cut-off threshold was picked for the best diagnostic performance to classify EAC/HGD using these recorded fluorescence videos. We introduce each step in more detail in this automated CAD part 2 as follows.

2.2.3.1 Patients' T/B ratios via Histogram-based method

Artifact frames with bright and sometimes colored boundaries are mainly caused by reflection from the optical lens and often result in high T/B ratios. A histogram method was adopted to remove this artifactual effect. T/B ratios of each fluorescence video were first represented as a histogram. Then frames with low or no fluorescence signals ('healthy' frames), which by design cannot be segmented using the Chan-Vese algorithm and have T/B ratio equal to 1.0, were removed from the histogram. Upper cut-offs derived from the histogram are applied to remove noise and artifactual effects (mentioned above). Since these video frames contain the same regional tissue that is eventually removed by EMR and assigned a single clinical evaluation, statistics (Mean and Median) of the histogram are applied to acquire the characteristic metrics of all the frames and to compute a single T/B ratio for the patient from all resulting video frames (Figure 2.2). This single value represents the T/B ratio for the patient. We tested different combinations of statistics (Mean and Median) and histogram cut-off (lower 50% ~100%) for comparison. The optimal combination was selected based on the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. For each combination of statistics and histogram cut-off, the

T/B ratios for all the 50 patients were obtained. By setting different diagnostic cut-off thresholds for the T/B ratio and comparing to the diagnoses based on the most severe histologic finding in the patient, we generated ROC curves. The combination with the largest AUC was selected for further experiments and evaluation.



Remove T/B ratio = 1.0

Calculate mean T/B ratio of low 90% frames

Figure 2.2 Histogram-based method to remove artifact frames. Frame with T/B ratio equal to 1.0 were first removed from the histogram, then a combination of statistics and histogram cut-off is applied for computing a single T/B ratio for the patient. In this example, this single value is calculated as mean T/B ratio of lower 90% of all frames.

2.2.3.2 Threshold for automatic diagnosis

We used the Youden index [39] of the ROC curve to select the optimal cut-off threshold for HGD and EAC classification. Youden index calculates the maximum vertical distance between ROC curve and the diagonal line. Normally, Youden Index = Sensitivity + Specificity - 1.

2.2.3.3 Classification evaluation

We evaluated our proposed method for automatic EAC/HGD classification using leave-one-out cross-validation (LOOCV). For each round of LOOCV, 49 videos were used as training set to decide the optimal cut-off using Youden index, which was then tested on the other remaining video. An average cut-off threshold of 50 rounds will be reported, as well as the classification accuracy, the average specificity and sensitivity over 50 cross-validation tests.

2.3 RESULTS

2.3.1 *Segmentation of ROIs with morphology operation*

All 50 fluorescence videos included in this study were able to be automatically segmented to identify the ROIs and to calculate T/B ratios. Each of the video frames was first smoothed with a 5*5 Gaussian filter and then segmented using the Chan-Vese algorithm. The parameters, shown as follows, were empirically taken as baseline, with $\mu=1$, $\nu=60$, $\lambda_1=2.0$, $\lambda_2=2.1$, dt=0.5, h=1. The selected parameters were optimized for low contrast "patchy" fluorescence images. In the developed user interface, the penalty term ν can be increased to limit area size inside the contour with a slider. Figure 2.3 illustrates segmentation (c) and morphology (d, e) results starting from an initial input video frame. Targets smaller than 10 pixels were removed by applying erosion (d). Neighboring contour with a width of 30 pixels is used as background to calculate T/B ratio (e). Red highlighting of the targeted boundary is exhibited in (f).

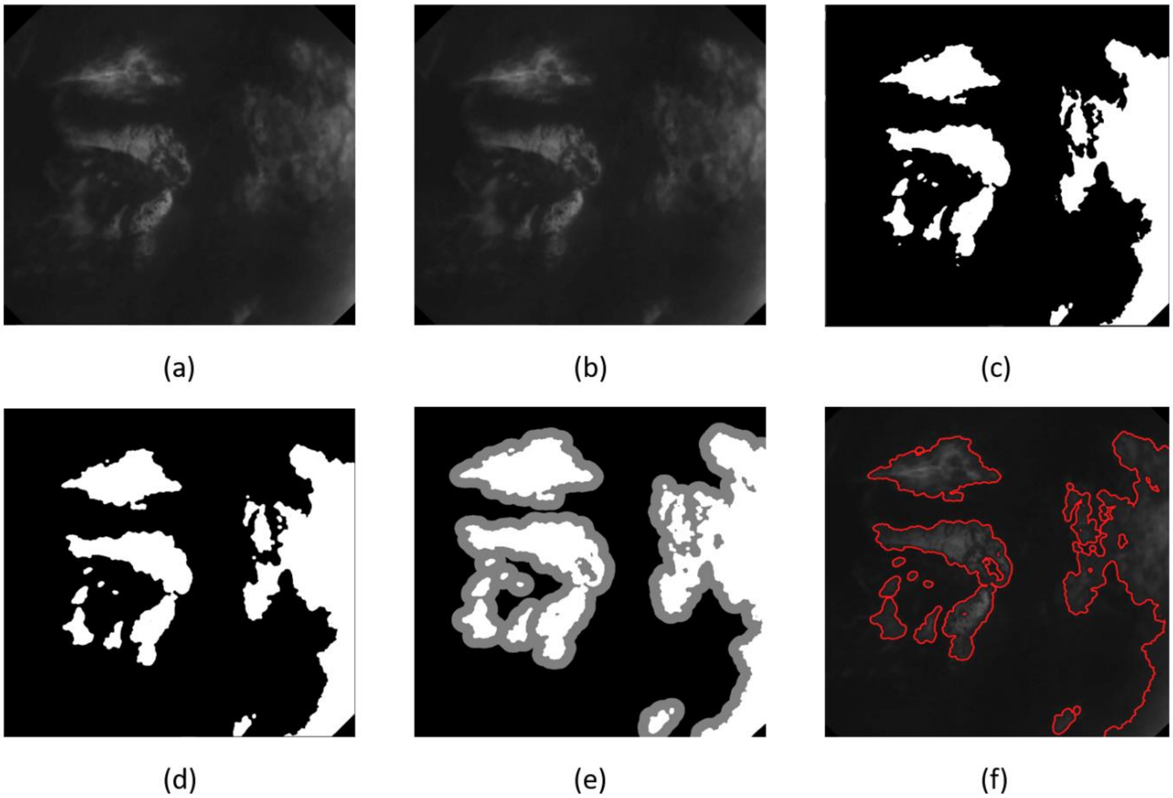


Figure 2.3 (a) Raw fluorescence frame. (b) Image after 5*5 kernel Gaussian filter. (c) Segmentation result with parameter $\mu=1$, $\nu=60$, $\lambda 1=2.0$, $\lambda 2=2.1$, $dt=0.5$, $h=1$. (d)Image after erosion applied to remove small isolated pixels. (e) Image after dilation to generate surrounding background band with a width of 30 pixels (grey band). (f) Highlighted fluorescence target ROI (inside red contour) with T/B ratio greater than the set threshold.

2.3.2 *T/B ratio calculation*

T/B ratios were computed from labeled ROIs. The following analyses were conducted to quantitatively evaluate the performance of our proposed approach.

2.3.2.1 Sensitivity analysis

We tested how target and background size effected T/B ratio.

Target size: Due to non-uniform target intensity in certain fluorescence images, changing the constraints on contour area size will change the boundary of the target, thus changing the T/B ratio.

Figure 2.4 shows an example that as we constrained the contour to brighter and smaller targets and maintained the remaining parameters unchanged, the T/B ratio would increase.

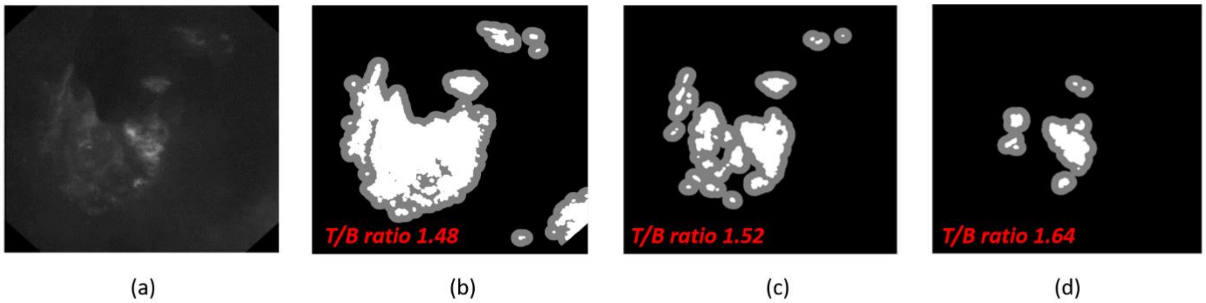


Figure 2.4 (a) Collected fluorescence image without processing. (b) T/B ratio 1.48, calculated with parameter μ=1, **<u>ν=60</u>**, λ1=2.0, λ2=2.1, dt=0.5, h=1, B=30 pixels. (c) T/B ratio 1.52, calculated with parameter μ=1, **<u>ν=180</u>**, λ1=2.0, λ2=2.1, dt=0.5, h=1, B=30 pixels. (d) T/B ratio 1.64, calculated with parameter μ=1, **<u>ν=330</u>**, λ1=2.0, λ2=2.1, dt=0.5, h=1, B=30 pixels. Comparing (b) ~ (d), with other parameters unchanged, more constraints were applied on the area size inside the contour, which resulted in smaller target area and increased T/B ratio.

Background size: We applied dilation to generate background band surrounding the tissue. As we include more pixels as background, the average intensity of background decreases, thus increasing T/B ratio (Figure 2.5), which is consistent to all fluorescence videos. This is because the background is non-uniform, with lower intensity further from the target. In this work, we regarded neighboring contour with a width of 30 pixels as background, which is the same standard as Joshi et al. [35].

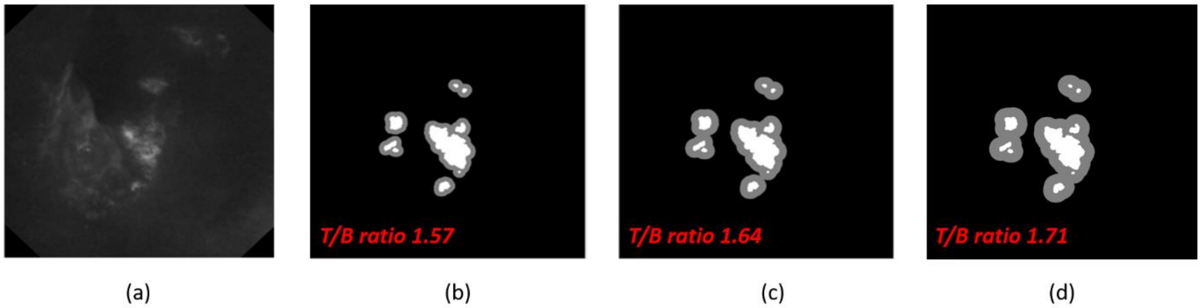


Figure 2.5 (a) Collected fluorescence image without processing. (b) T/B ratio calculated using surrounding background with a width of (b) B=20 pixels, T/B=1.57 (c) 30 pixels, T/B=1.64 and (d) 40 pixels, T/B=1.71. Comparing (b) ~ (d), with other parameters unchanged, wider surrounding background resulted in higher T/B ratio.

2.3.2.2 Speed performance

The total processing procedure is implemented in C++ using openCV library. The bulk computation time for calculating T/B ratio for each frame is from the segmentation procedure. For each evolution of the level set function, the computing speed is almost constant. However, to get a reliable segmentation (repeatability >98%), maximum of 8 iterations was used. The average processing time for each frame was 0.5s. The speed was measured using an Intel Core i7-4980HQ @ 2.80GHz-CPU PC with 16 GB memory.

2.3.3 *Histogram based method performance*

AUC is used to select the optimal combination of statistics and histogram cut-off for computing a single T/B ratio for each patient. Based on the T/B ratios for all 50 patients, we calculated the sensitivity and specificity by setting different diagnostic T/B cut-off thresholds, generated ROC curves and obtained AUCs. AUCs from combinations of statistics and histogram cut-offs (lower 50% ~ 100%) are shown in Figure 2.6. By selecting the maximal AUC of 0.875, the optimal way to compute T/B ratio for each video was determined to be the median of lower 95% frame T/B ratios after removing frames with T/B ratio equivalent to 1.0. Compared to mean, the median is less susceptible to outliers and extreme values. Differences between AUCs using mean versus median statistics and using histogram cut-off of lower 50%-100% were not statistically significant, although the sample size of this study limits power to detect small differences.

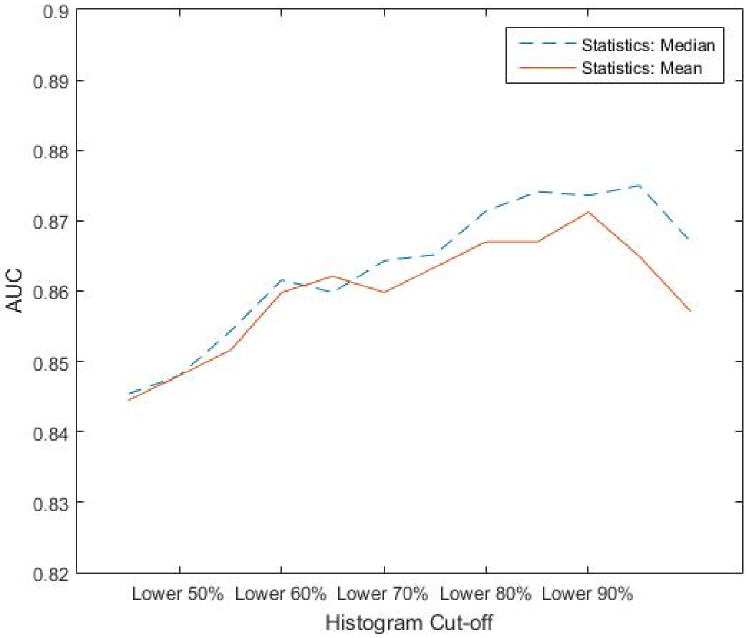


Figure 2.6 AUC from different statistics and histogram cut-offs to calculate T/B ratio for patients. The best combination to compute T/B ratio for each video is median of lower 95% frame T/B ratios after removing frames with T/B ratio equivalent to 1.0, with AUC 0.875.

2.3.4 *Classification evaluation*

For each patient, T/B ratio was calculated using the best combination of statistics and histogram cut-off selected above and classified by the highest grade found on pathology (Figure 2.7a). The mean T/B ratio, as well as variance for each category is plotted in Figure 2.7b. As we can clearly see, mean T/B ratios for either EAC or HGD were greater than those for BE, gastroesophageal junction (GEJ) and LGD. However, the mean T/B ratio for squamous (SQ), which is the normal case, is larger than BE and GEJ. This is because there are only 2 SQ cases and one false positive case that is retained in the analysis appeared to bias these results Different diagnostic cut-off T/B thresholds were set and sensitivity and specificity curves were generated using T/B ratios for all 50 patients (Figure 2.7c). At a target/background ratio of 1.33, which was calculated by Youden index, there was 94% sensitivity and 75% specificity. At a target/background ratio of 1.45, the specificity was 88%, while the sensitivity was 66%. The ROC curve comparing the T/B ratios for early neoplasia (HGD and EAC) with those for the other classifications (SQ, BE, LGD, and GEJ) achieved an AUC of 0.875. LOOCV was performed to evaluate our proposed method for automatic EAC/HGD classification, which prevents overfitting and combines average of classification error to derive a more accurate estimate of classification performance. The average cut-off threshold calculated from LOOCV is 1.34. The average sensitivity and specificity are 0.92 and 0.75, respectively. The average AUC of ROC curve is 0.871. With 43/50 accurate classification on one remaining test set, the classification accuracy is 86%. These results are used as the classifier results, with 1.43 as cut-off threshold for highlighting ROIs for biopsy and endoscopic therapy. The sensitivity and specificity for automatic EAC/HGD classification method is 0.92 and 0.75, respectively.

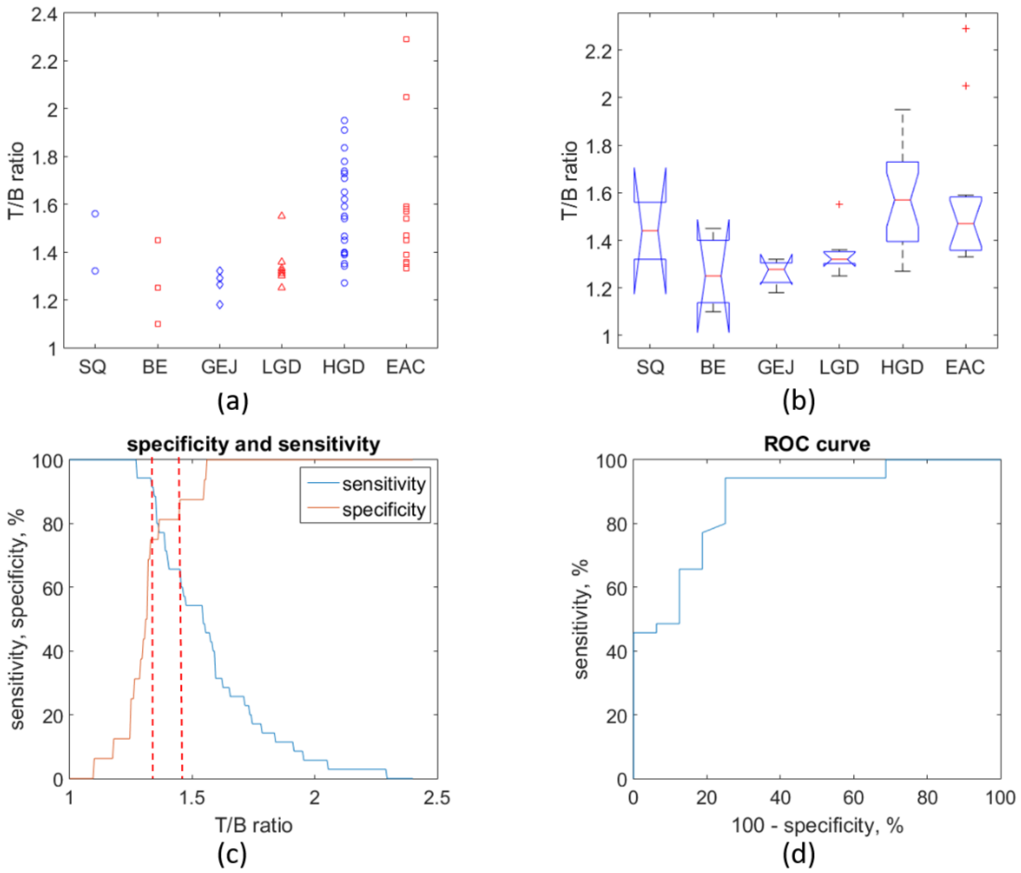


Figure 2.7 (a) T/B ratios for individual patients grouped by the most advanced grade found on pathology (b)The mean and standard deviation values for T/B ratio in each category. The mean values were 1.44 for SQ (n=2), 1.27 for BE (n =3), 1.26 for GEJ (n =4), 1.35 for LGD (n =7), 1.57 for HGD (n =21), and 1.56 for EAC (n =13) (c) At a target/background ratio of 1.33, sensitivity was 94% and specificity was 75%. At a target/background ratio of 1.45, specificity was 88%, and sensitivity was 66%. These are the results calculated using T/B ratios for all 50 patients. The final results will be calculated as the average from LOOCV (see above) (d) The ROC curve comparing the T/B ratios for early neoplasia (HGD and EAC) with those for the other classifications (SQ, BE, LGD, and GEJ) had an AUC of 0.875.

2.4 DISCUSSION

In this chapter, we demonstrated a procedure to calculate T/B ratio in quasi-real-time using human clinical data that takes advantage of a novel fluorescence peptide and a custom endoscope used in vivo. Relative quantification was provided by the amount of molecular probe binding in the target region compared to the surrounding tissue (background), based on the relative fluorescence intensity that is measured from both regions within a video frame. T/B ratios above a threshold are indicative of progression to EAC from Joshi's previous work [32]. We extend the use of T/B quantification to provide potential real-time highlighting of targeted regions that are above a set threshold. This set threshold is calculated using our automatic CAD algorithm (part 2) for classifying cancer (EAC) and its precursor (HGD). LOOCV is performed to evaluate this method and to prevent overfitting. The average T/B threshold from LOOCV, 1.34, is used as cut-off threshold for highlighting ROIs for biopsy and endoscopic therapy. A user-interface is designed to highlight the targets by enhancing the color of the targets or simply encircling the targets with red boundary and the latter was chosen by the co-authoring gastroenterologists (Figure 2.8). A toggle switch is being designed to change the cut-off threshold depending on the clinical scenario, routine surveillance (biopsies) or therapy (EMR). The design iterations are continuing for this clinical interface.

The initial processing speed is 2fps, while the ideal processing speed is to match the frame rate of the endoscopy, which is above 5 fps. The processing speed can be accelerated using graphics processing unit (GPU) parallel processing. This processing speed can be further accelerated by using field-programmable gate array (FPGA) architecture, which is available in our developing scanning fiber endoscope (SFE see below). The most time-consuming part of the processing is the segmentation step. Instead of processing the iterative evolution of the level set function with only one processor, FPGA can pipeline each iteration with multiple processors. In addition, using FPGA parallel processing makes the acceleration possible. This method demonstration is designed to help an experienced endoscopist to locate and track suspected targets measured to be the highest risk for cancer for subsequent biopsy or surgical removal.

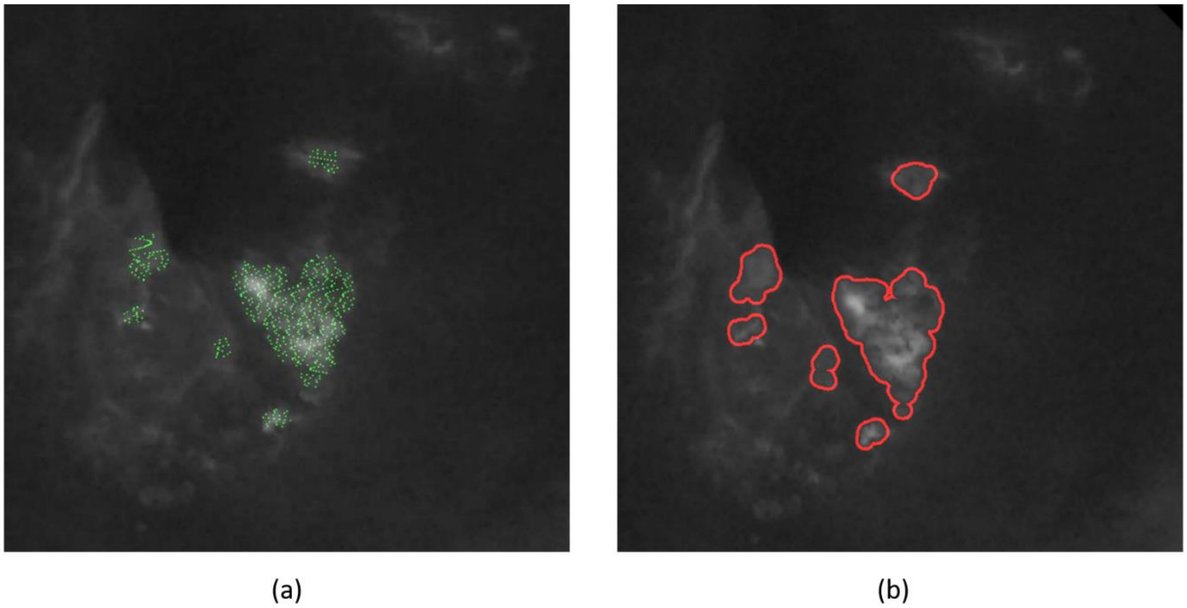


Figure 2.8 Two methods to highlight the targets. (a) Enhancing the color of the targets. (b) Encircling the targets.

We developed an automatic CAD algorithm for classifying esophageal neoplasia (part 2), using the same recorded endoscopic video frames from Joshi et al. [35] that captured fluorescence from targeted tissues being removed by EMR or biopsy, and confirmed by histopathology. This unique data set consists of a series of video frames from the same region of targeted tissue that is eventually removed and assigned a single clinical evaluation. By analyzing the series of video frames for each case, an overall diagnostic T/B value can be assigned for each patient. In calculating this overall diagnostic T/B value, repeated measures of the T/B ratio are made for each video frame. This allows a sensitivity analysis under clinical conditions of a range of variables, which test the robustness of our image processing algorithms. Analysis on surrounding background has been applied to test the sensitivity. When including more pixels as background, T/B ratio increases. To mitigate the effect of background size on T/B ratio, we regarded neighboring contour with a width of 30 pixels to match Joshi et al. [35] post-processing study results, which are compared below (Table 2.1) to our automated CAD approach. The future clinical utility of this method is expected to aid a less experienced endoscopist in the referral decision making process for suspicion of EAC.

Table 2.1 Comparison between mean T/B ratio for individual patients grouped by histopathology classification

	SQ (n=2)	**BE (n=3)**	**GEJ (n=4)**	**LGD (n=7)**	**HGD (n=21)**	**EAC (n=13)**
Joshi's results	1.21	1.39	1.27	1.39	1.73	1.65
Our results	1.44	1.27	1.26	1.35	1.57	1.56
Difference in %	19.0%	-8.6%	-0.8%	-2.9%	-9.2%	-5.5%

SQ: squamous, BE: Barrett's esophagus, GEJ: gastroesophageal junction, LGD: low-grade dysplasia, HGD: high-grade dysplasia, EAC: esophageal adenocarcinoma

We can see in Table 1 that our overall mean T/B ratios for each category are smaller than those for Joshi's, except for the SQ case. The variables or clinical issues that make this difference in T/B ratio between our automated method and Joshi's method using the same video data set assume to be (1) different segmentation algorithms to label fluorescent targets (2) user selection of video frames by Joshi et al and (3) inconsistent distance between the detector and the imaged mucosa [32]. The Otsu's segmentation algorithm that Joshi applied, is limited by excluding small targets with lower intensity from the labeled target region, since this algorithm applies global threshold to segment the image. Instead, we applied Chan-Vese algorithm, which is more robust to noisy, low contrast images. Our segmentation algorithm includes more small targets with lower intensity, thus decreasing T/B ratio. Moreover, Joshi selected video frames with high quality for post-processing, while we calculated each frame in the videos. So, there is no interobserver variation in selecting the frames in our method, and there will be some systematic differences in original content when comparing to Joshi et al. Finally, to compensate for uncontrolled or inconsistent distance between the detector and the imaged mucosa, Joshi et al., [35] generated a ratiometric image (fluorescence image/reflectance image) before calculating the T/B ratio. However, since the fluorescence image and the reflectance image are not captured concurrently and there is registration artifact when performing imaging procedure in time series, it is impossible to register all frames in the fluorescence video with the reflectance video. The ratio image suppresses the noise and increases the contrast of the image thus increasing the T/B ratio. However, our mean T/B ratio for SQ are 19% greater than Joshi' results. The main reason assumes to be small sample size of the SQ cases and one false positive case among the 2 SQ cases biases the results.

Table 2.2 Comparison between results from our automated CAD method and Joshi's post-processing CAD method

	T/B Cut-off threshold	**Sensitivity**	**Specificity**
Joshi's results	1.49	76%	94%
Our results	1.34	92%	75%
PIVI recommendations	---	90%	80%

At a T/B ratio of 1.34, the average sensitivity and specificity from LOOCV using our method are 92% and 75% separately, compared to Joshi's of 76% and 94% at T/B ratio at 1.49, respectively (described in Table 2.2). Even though we have different cut-off threshold, sensitivity and specificity compared to Joshi's, we returned a similar AUC for the ROC curve, 0.871 compared to Joshi's 0.884 [32], which indicates both methods can be applied to calculate T/B ratio for patients. The difference in sensitivity and specificity lies in different calculated T/B ratios, different selected cut-off threshold and systematic difference. In this paper, the cut-off threshold for the ROC curve was calculated using Youden index. However, in practice, the cut-off threshold should be incorporated with the clinical scenario for correct and false diagnosis. For example, when taking routine surveillance and guiding biopsies, the cut-off threshold should obtain a high sensitivity to rule out the possibility of false negative. When performing endoscopic resection, a procedure with greater risks to the patient than biopsies, the cut-off threshold should generate a higher specificity to decrease the possibility of false positive. Overall, we have reasonable preliminary results for a future EAC surveillance program, compared to Joshi's. As stated in the PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) report, the new imaging technology should have a recommended sensitivity of 90% or greater and maintain a specificity at least 80% for detecting HGD and EAC during the endoscopic surveillance of patients with BE [40]. Our automated CAD for EAC and HGD met the criterion for sensitivity at 92% but failed to meet the criterion for specificity at 75% in this preliminary study with 50 BE subjects.

One of the uncontrolled variables in our approach lies in the unfixed segmented boundary. Therefore, change in the penalty on the area inside the contour will result in change in segmented target size, thus changing the T/B ratio in some frames as shown in 2.2.1. Clinical tests should be conducted to evaluate the expected range of target sizes of HGD lesions for EAC which may be different for sub-populations and of course other cancers. Another uncontrolled variable is that there is no distance compensation in our approach, which can assist in removing artifacts and noise in the image and enhancing image contrast. Moreover, we only applied a single fluorescence peptide probe to detect one biomarker. The last but not least, we calculate the average T/B ratio of multiple targets to represent the T/B ratio of the frame. If the intensity for the targets is not uniform, then the average T/B ratio underestimates the brighter targets. However, this issue can be resolved by the endoscopist adjusting the position of the scope closer to the brighter targeted tissue, as our method has the potential to recalculate T/B ratio in real time under these dynamic conditions.

To overcome some of these limitations, we are developing the SFE for first-in-human multimodal endoscopic procedures using multiple fluorescence peptides from Dr. TD Wang's group at the University of Michigan (discussed in chapter 4). Since the SFE has an ultrathin and flexible profile, so there can be a range of different manifestation to reach the esophagus, such as a mini-endoscope within the working channel of a conventional endoscope or as a transnasal design or tethered-capsule design [41]. Compared to the much larger caliber Olympus system, our SFE can acquire fluorescence and reflectance videos concurrently, which can account for the inconsistent distance between the detector and the imaged mucosa [42]. Furthermore, other systematic errors can be reduced with multimodal SFE imaging, such as real-time autofluorescence background mitigation algorithm and elimination of specular reflection artifacts, thus increasing accuracy and scale of the T/B value by over an order of magnitude [43, 44]. These improvements are expected to increase specificity to reach the desired minimum of 80% (PIVI Recommendations, Table 2) while maintaining automation of the analysis. Moreover, the SFE has multiple channels in a wide range of detection wavelengths (visible to infrared), which makes multiple molecular targets detection possible. The last but not least, the FPGA architecture in the SFE makes real-time guiding biopsy using multiple fluorescence channels possible. Multiple molecular probe imaging of early cancer is expected to increase sensitivity without sacrificing specificity due to the high molecular specificity of the peptide fluorescence probes [45]. Comparison between Joshi's method w/Olympus, our method w/Olympus, and future work in esophagus w/SFE is described in Table 2.3 below.

Table 2.3 Joshi's method w/Olympus, our method w/Olympus, and future work in esophagus w/SFE

	Olympus + Joshi's Method	**Olympus + Our method**	**SFE + Our method**
CAD with T/B ratio	Yes	Yes & Automatic	Yes & Automatic
Segmentation algorithm for T/B calculation	Otsu's algorithm: Single threshold	Chan-Vese: Active contour, robust to patchy, noisy images	Chan-Vese: Active contour, robust to patchy, noisy images
Video frame selection	Manual	All	All
Fluorescence & reflectance capture	Time series	Time series	Concurrent
Distance compensation	Yes	No	Yes
Multiple fluorescence channels	No	No	Yes
Real-time capability	Post processing	2 fps	5-30fps

In our future work using either multimodal endoscopy system, we aim to calculate T/B ratio in real time. However, if the FPGA accelerated T/B calculation using Chan-Vese algorithm cannot match the frame rate of the endoscopy, deep learning based T/B calculation will be investigated. Compared to Chan-Vese algorithm, deep neural network architectures, such as UNet, show promising results in real-time image segmentation [123]. In addition, the networks can learn features from both high and low contrast images and show potentials in segmenting "patchy" fluorescence images. In the chapter 5, the feasibility of real-time T/B calculation using deep learning algorithms will be demonstrated. The performance of the two methods, Chan-Vese and deep learning, will be compared. The combination of SFE and automatic, real-time T/B calculation is estimated to greatly improve the efficiency and safety of guiding biopsy and surgical removal of neoplasia in the esophagus and smaller ductal regions.

Chapter 3. CALIBRATION OF NIR FLUORESCENCE IMAGING FOR TUMOR SURGICAL MARGIN DELINEATION: MULTI-STEP REGISTRATION OF NIR FLUORESCENCE AND HISTOLOGICAL IMAGES

In this chapter, I describe a novel framework to register NIR fluorescence images to histological images, thus evaluating fluorescent probes in delineating tumor cores and margins. This pipeline involves distance compensation, specular/ambient light rejection from fluorescence images, a multi-step registration between histological slices and calibrated fluorescence images, which is then validated using a brain tumor mouse model.

3.1 BACKGROUND

Surgical resections are often the primary treatment option for both benign and malignant tumors. Curative surgery for many cancers requires a wide local excision to achieve negative margins [46, 47]. Although a greater extent of tumor resection is critical for reducing tumor recurrence and improving prognosis, wide negative margins for brain surgery are not possible [48, 49, 50]. At the margins, the delicacy and unique functionality of the surrounding healthy tissue necessitates extreme precision during excision for neurosurgery, as the risk of damaging normal brain tissue is the impairment of patient's cognitive function for their lifetime. In intraoperative neurosurgery, the ability to assess tumor margins with high accuracy would benefit patient's survival while preserve quality of life after surgery.

Visual appearance and palpation have been the standard methods used by surgeons to differentiate tumor from healthy tissue. These techniques are ineffective at the margins for many tumors, where high contrast and resolution are required intraoperatively. Bright tumor specific indicators, such as targeted fluorophore, can visually enhance tumor margins, which is ideal for residual tumor clean-up and removal of disturbed tumor fragments. Compared to other noninvasive imaging technologies, such as MRI, CT, PET, and ultrasound, advanced optical imaging is the most pragmatic means to provide high-resolution, video-rate, and non-contact intraoperative imaging for surgeons.

In the past 30 years, various optical imaging techniques have been investigated to guide surgery and to delineate tumor margins [51-57], which can be categorized into three mechanisms: (1) structural contrast enhancement, (2) metabolic contrast enhancement, and (3) molecular contrast enhancement.

(1) Optical coherence tomography (OCT) is a non-contact imaging modality capable of detecting photons backscattered from tissue with high sensitivity. The clear advantage of applying OCT in surgical procedures is to provide three-dimensional tissue images, which may benefit tumor resections with additional sub-surface anatomy visualization [58]. However, OCT provides structural data without molecular or chemical contrast, so that enhanced contrast of tumor cells is not well developed.

    Hyperspectral imaging (HSI) is a non-contact, label free imaging technique that can acquire a wide range of optical spectral information compared to traditional 3-channel (RGB) images. By extracting information from a greater spectral range with finer resolution, HSI can clearly delineate tumor tissue from healthy brain tissue [53]. However, the computational task and processing time impedes HSI as a real-time intraoperative imaging technique.

Nontargeted fluorophores, such as fluorescein, methylene blue (MB) and indocyanine green (ICG), have been investigated to enhance tissue and vessel structures, especially in tumor regions, where these contrast agents extravasate passively out of the vasculature and accumulate. This phenomenon is commonly referred as enhanced permeability and retention (EPR) effect. Fluorescein has been applied for guiding high-grade glioma resection using a surgical microscope with excitation and emission filters and appears safe and effective in clinical trials [59]. MB is another clinically approved dye with visible and near infrared (NIR) fluorescence. Although MB NIR fluorescent property is weak and inefficient, MB has been proposed for fluorescence-guided interventions using a hand-held device [60]. In contrast, ICG is a high-contrast NIR fluorescence indicator that has been widely used to guide surgical resections in humans. The clear advantage of ICG is the excitation well beyond the visible spectrum at >780 nm, which produces negligible autofluorescence and excites targets located more deeply from the surface of the tissue. By combining non-specific ICG and high-resolution confocal endomicroscopy, neurosurgeons at the Barrow Neurological Institute have visualized brain tumors and their margins [51]. However, endomicroscopy is limited to small field of view (FOV) and often requires tissue contact for a robust measurement. Without conjugation to a targeting moiety that accumulates in cancer, the results of using the pure fluorescence dye have been inconsistent and lack specificity [61].

(2) Metabolic contrast refers to a contrast agent that is activated or converted from a non-fluorescent to a fluorescent form due to a metabolic process. 5-aminolevulinic acid (5-ALA) is the most widely investigated contrast agent for guiding surgery, which has been recently approved by the US FDA as an intraoperative imaging agent for the resection of

malignant glioma [62]. In cells with high metabolic activity such as tumor cells, 5-ALA is processed via the haeme-synthesis pathway to produce protoporphyrin IX (PpIX), which emits fluorescence in red visible light range (635 nm peak) and can provide direct visualization of the tumor [63]. By combining PpIX with highly efficient fluorescence cameras, neurosurgeons can identify tumor margins with low PpIX concentrations, which are not visible with standard surgical microscope systems [64]. However, the shallow tissue penetration of the PpIX visible light excitation and emission limits subsurface tumor delineation.

(3) Molecular imaging is an emerging technique that can provide molecular biology information specific to cancer progression. Targeting on the differences in tumor molecular expression, new agents, especially in the NIR range, are in various stages of preclinical or clinical investigation, including labeled antibodies, affibodies, peptides and other molecules [65, 66, 35]. The combination of NIR targeted fluorescent probes with advanced wide-field fluorescence imaging devices has contributed to a surge of interest in the field of fluorescence-guided surgery [8]. Recently a chlorotoxin (CTX) has been conjugated to a derivative of ICG that has specific binding affinity to cancers such as induced gliomas in mouse models [67] and spontaneous soft tissue sarcomas in canine models [68]. Bevacizumab-800CW, a monoclonal antibody to vascular endothelial growth factor, has been investigated for the diagnosis of esophageal adenocarcinoma [57]. Rosenthal et al. used cetuximab conjugated to IRDye800 for surgical navigation in head and neck cancer [65]. The clear advantage of using NIR fluorophore is their improved signal to noise at the margins due to reduced autofluorescence and tissue optical scattering compared to visible light fluorophores.

To verify tumor margins detected by fluorescence images, the in vivo imaging must be compared with pathology which provides the ground truth of cancer regions. The different tissue appearances from the two imaging modalities makes this multimodal registration task challenging. Moreover, variations in tissue optical properties and tumor depth makes correlating fluorescence images to histopathological slides more difficult. Elliott et al. utilized IRDye800CW labeled affibody specific to epidermal growth factor receptor to guide glioma resection [69]. The tumor was correlated to fluorescence contrast using a multistep image registration. However, in their approach, fluorescence images were directly registered to histological slides. Structural landmarks in fluorescence images were selected as control points for registration, which are fewer and less obvious compared to those in reflectance images. Lu et al. applied principle component analysis to extract macro images from HSI and registered these extracted macro images to histological images, thus evaluating HSI as an intraoperative visual aid for surgical resection [70]. However, both registration methods compared to a single-layer histological slide, which was selected subjectively and might not contain complete tumor information.

In this chapter, a multi-step registration framework is established to correlate fluorescence images to stacked histological images, which can be applied to evaluate fluorescent probes in delineating tumor cores and margins. Since variable experimental parameters and artifacts may affect accurate evaluation of fluorescence imaging, a series of preprocessing steps are applied to calibrate fluorescence imaging for improved accuracy and standardization, as stated in [14]. An optical attenuation model is applied to integrate multiple histological slides and project deeper tumor regions to the surface. Preliminary aspects of this model were presented in Image-Guided Procedures, Robotic Interventions, and Modeling panel within 2018 SPIE Medical Imaging [56].

The whole framework is verified using a mouse model of brain tumor labeled with NIR molecular probe, BLZ-100 (Blaze Bioscience Inc., Seattle, WA) [67]. Multimodal scanning fiber endoscope (mmSFE), a wide-field multimodal imaging device that captures fluorescence and reflectance simultaneously with exact spatial registration, is used to image BLZ-100 tumor indicator ex vivo.

3.2 MATERIALS AND METHOD

3.2.1 *Mouse brain tumor model*

A patient-derived orthotopic xenograft medulloblastoma tumor model (Med-114FH) was used for brain tumor fluorescence imaging under the IACUC approved protocol at the Fred Hutchinson Cancer Research Center (IR no. 1457) [71]. Ten adult male and female NSG mice between 3 and 6 months of age were used in this study. Recipient mice were administered Buprenorphine SR analgesia by subcutaneous injection approximately 1 hour before the orthotopic implant. At the time of surgery, mice were anesthetized using inhaled Isoflurane. Establishment of anesthesia was determined by monitoring respiratory rate and toe pinch reflex. Under anesthesia, an incision of ~1 cm was made along the mediolateral line using a disposable scalpel and the skull was exposed and cleaned of any minor connective tissues or blood using cotton tipped applicators. Once the area was clean, a handheld microdrill (Ideal Microdrill #67-1000, Cell Point Scientific) with a 0.9 mm burr (19007-09 Fine Scientific Tools) was used to create a burr hole into the cortex. 2 µl of cell suspension (100,000 cells) was then deposited into the intracranial space by inserting a 10 µl Rainin Pipette-lite fitted with a 2-10 µl ART tips barrier non-filtered pipette tip (2139, Thermo Scientific). A small piece of surgical foam (09-0396-05 Pfizer Injectables) was placed onto the burr hole and the incision closed with veterinary-grade surgical glue (Vetbond, 3M). Following closure of the incision, mice were removed from Isofluorane and transferred to a clean recovery cage placed on a heat mat for duration of recovery. Mice were monitored regularly for signs of tumor formation, which can include distention of the calvarium, head tilt, reduced feeding, weight loss, dehydration, hunched posture, or poor grooming habits. Brain tumor symptoms were typically observed 4-6 weeks after implantation.

3.2.2 *BLZ-100*

BLZ-100 is composed of a modified form of the CTX peptide conjugated to the NIR fluorescent dye ICG, which has been shown to bind specifically to many types of tumor cells, including glioma and other brain tumors, breast, head and neck cancers, and sarcomas [67, 68, 72]. Assessment to tumor burden is made by trained researchers who use professional judgment and experience to determine that mice have clinically evident tumor burden but are not moribund. Upon visual identification of the tumor burden, mice were administered 10 nmol of BLZ-100 in 100 µl PBS by tail vein injection. A 30G needle attached to an insulin syringe was inserted into the right lateral tail vein of a mouse and the compound was administered as a single bolus injection. Mice were returned to the home cage with free movement and access to food and water for the 4 hours incubation period. After 4 hours of circulation, brains were harvested. The collected brains were stored in ice-cold Phosphate-buffered saline (PBS) buffer, shielded from light and transferred for mmSFE imaging.

3.2.3 *mmSFE imaging*

The mmSFE imaging system is specifically designed for wide FOV, high quality, video rate (30 Hz), concurrent multiple fluorescence imaging [76, 84]. In this study, a 4-channel endoscope was used for BLZ-100 imaging, one for reflectance (638 nm) and three for fluorescence including one NIR channel.

The whole brain surface was first imaged using the mmSFE at 1 cm away from the scope distal tip, which was then sectioned into 2 mm slices coronally. Tissue sectioning was performed using the sharpest instruments possible to allow for clean cutting and prevent damage associated with crushing or tearing of tissue. Double Edge Blades with thickness 0.10 mm (Electron Microscopy Sciences Cat. # 72003-01) were used to cut the brain that was 3 to 10 hours post euthanasia. Cutting blocks were used for high quality sectioning (Braintree Scientific 1 mm Coronal). Immediately after sectioning, the sliced brains were placed in ice-cold PBS buffer and imaged using mmSFE at 1 cm away from the scope tip. Reflectance and fluorescence images of the slices were captured with dimmed laboratory lights on. In summary, five slices were imaged using mmSFE for each mouse.

3.2.4 *Histology*

After mmSFE imaging, the 2 mm brain slices were fixed in 10% Neutral Buffered Formalin solution for 48 hours. Then routine hematoxylin and eosin (H&E) staining was done for histological evaluation. The formalin-fixed tissues were embedded in paraffin, which were later cut into 4 um thickness slides coronally at 100 μm intervals. For each tissue block, 5~6 slides were cut. The cut slides were stained with H&E and imaged using a Leica color pathology slide scanner (Aperio Scanscope AT, Leica Biosystems Imaging, Inc). The whole protocol for mmSFE imaging and histology processing is shown in Figure 3.1.

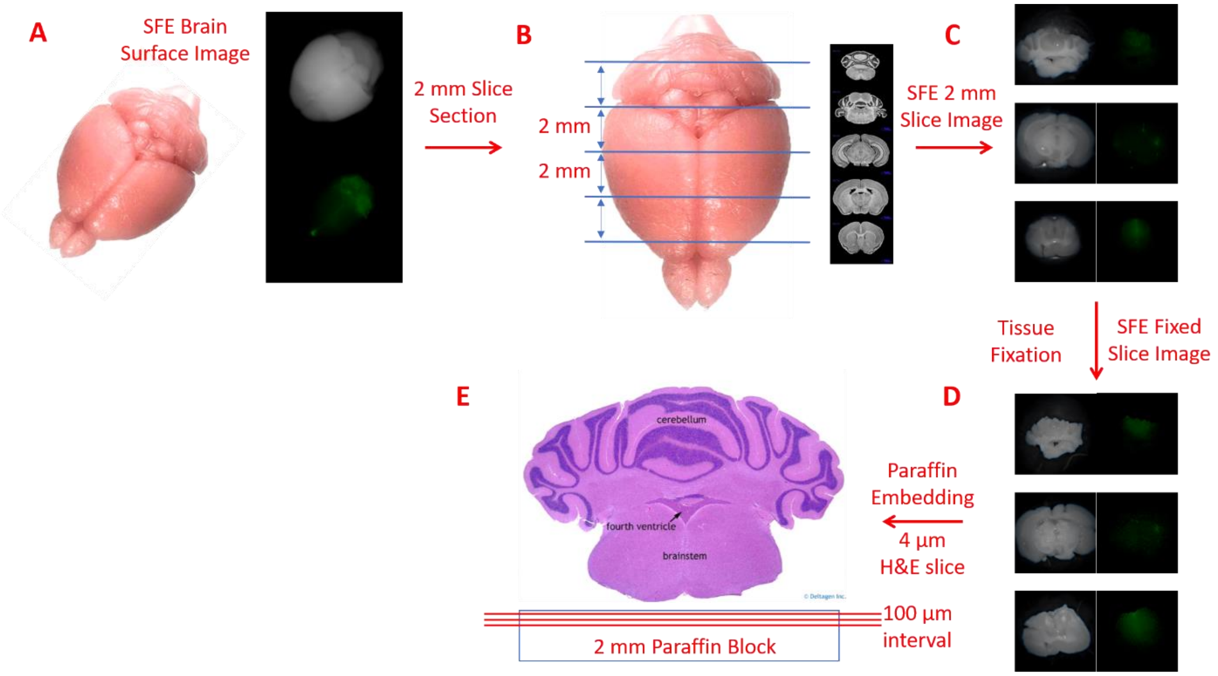


Figure 3.1 Protocol for mmSFE imaging and histology processing: The whole brain is imaged using mmSFE and then sectioned into 2 mm slices coronally. The sectioned slices are imaged using mmSFE. Routine H&E staining is performed on each 2 mm slice.

3.2.5 *Calibration of NIR fluorescence imaging for Tumor margin delineation*

3.2.5.1 Preprocessing methods for mmSFE fluorescence image

The fluorescence data acquired using mmSFE were preprocessed for accurate evaluation and standardization, which consisted of two steps shown below.

Specular reflection detection and removal

Specular reflections are usually observed from moist tissue surface, which can be a problem in minimally-invasive surgeries, where movement of the scope is limited, and the reflections cannot be moved off regions of interest. To provide quantitative evaluation of fluorescence imaging, these undesired reflections need to be removed. Figure 3.2 shows the pipeline to remove specular reflection. Specular pixels in reflectance channel were first detected using Matlab function multithresh, which applies Otsu's algorithm to segment specular reflection. A dilated specular mask was then generated using Matlab function imdilate with a 3*3 square structuring element. Since mmSFE captured co-registered reflectance and fluorescence images, the generated specular mask from reflectance would apply to both reflectance and fluorescence channels. Each pixel in dilated specular mask was then replaced by the average of neighbor pixels without specular reflection for both reflectance and fluorescence images. A neighbor size 7*7 was applied. If the neighbor pixels were all affected by specular reflection, we increased the neighbor size until there were pixels without specular reflection.

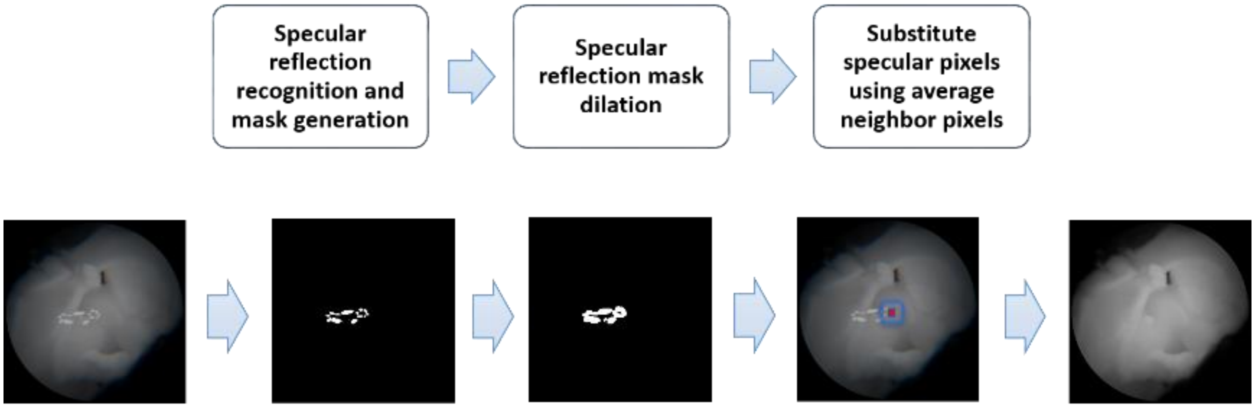
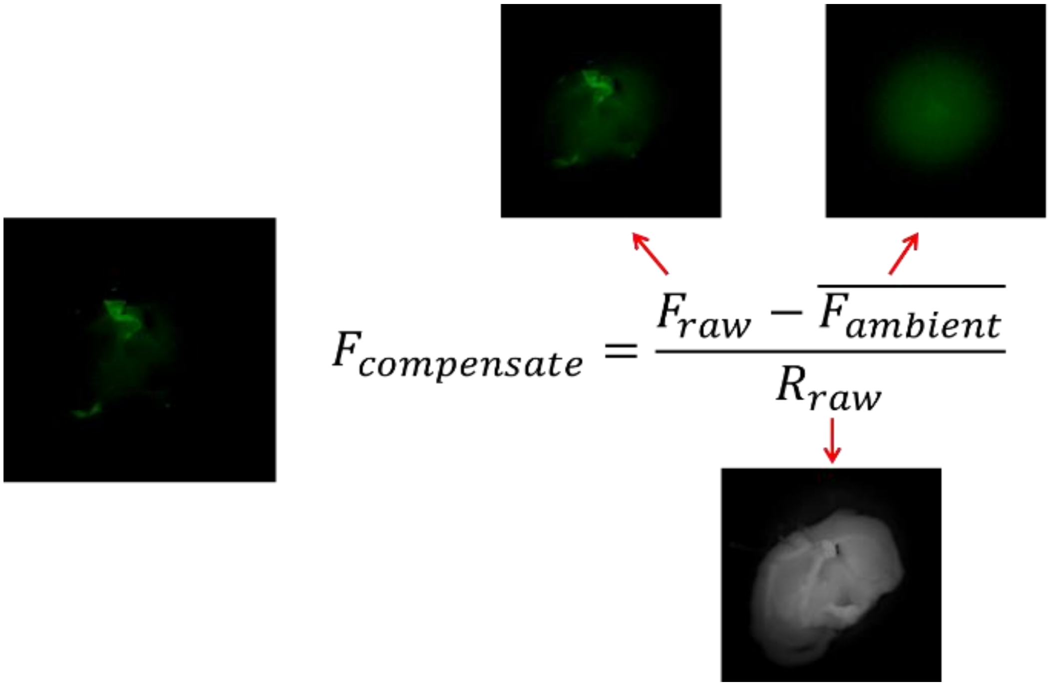


Figure 3.2 Pipeline for specular reflection removal: Specular reflection regions are first detected, and a dilated specular mask is generated. Each pixel in the dilated specular mask is then replaced by the average of neighbor pixels.

Fluorescence normalization

The fluorescence background was previously measured by placing a non-fluorescent white reference 1 cm away from the scope tip with the NIR excitation source, which was later subtracted from raw fluorescence to reject ambient light. To compensate for inconsistent distance between the scope distal tip and various points on the imaged tissue, a ratiometric method was applied. From any point on the tissue surface, the fluorescence and reflectance generated will travel along identical paths to the scope end face. Therefore, the difference in distance can be compensated using the ratio between these two images. Since mmSFE can capture reflectance and fluorescence images concurrently, there is no registration error between these two channels, which makes this ratiometric method more robust. The method was first tested using a uniform gelatin phantom (10%), mixed with ICG. Gelatin powder (G2500, Sigma-Aldrich) is mixed with distiller warm water (40 °C). The gelatin/water mixture is stirred until all the gelatin is fully dissolved, resembling a uniform liquid. ICG (1340009, Sigma-Aldrich) solution is then dispersed in the mixture. The final concentrations of ICG is 1 μM. An example for this normalization process is shown in Figure 3.3. To avoid divide by 0 errors, each pixel in the reflectance images is added by 1.



$$F_{compensate} = \frac{F_{raw} - \overline{F_{ambient}}}{R_{raw}}$$

Figure 3.3 Ambient light rejection and distance compensation (F-fluorescence, R-reflectance): Ambient light is first subtracted from the raw fluorescence. A ratiometric method is applied to compensate for inconsistent distance between the scope distal tip and various points on the imaged tissue.

3.2.5.2 Multi-step fluorescence registration & 2D tumor map synthesis

A multi-step image registration procedure was developed and applied to the preprocessed fluorescence image to enable pixelwise comparisons with stacked histological slides. A surface macro histological image was first synthesized from 3D histological slide deck. Then a multi-step registration is conducted between mmSFE reflectance image and the surface macro histological image. The transformations calculated from reflectance image were applied to the corresponding fluorescence image. A 2D tumor map was extracted from histological slides using optical attenuation model and average intensity projection (AIP), which was used as ground truth for comparison. The detailed procedure follows the steps described below.

Surface macro histological image synthesis

NIR light penetrates deeper in tissue, which excites fluorescence beneath tissue surface. To better correlate fluorescence with histological images from both surface and deeper tissue, a 3D histological slide deck was first generated. Each slide in the 3D deck was registered to its adjacent slide using Matlab functions *imregtform* and *imregdemons*, which perform affine transformation and non-rigid registration, respectively. The top-view surface pixels were first extracted from the 3D histological slide deck and projected to a 2D plane to synthesize a surface macro histological image, which was then used to register with mmSFE reflectance image. This surface macro histological image provided compensation for some incomplete histological slides being cut from uneven tissue surface in paraffin blocks (Figure 3.4).

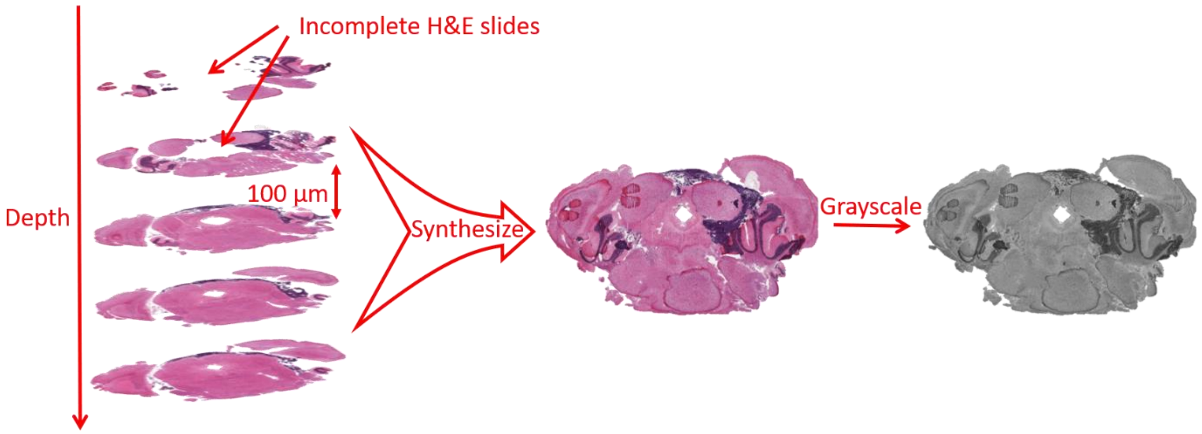


Figure 3.4 Synthesized macro histological image: Surface projection is applied to synthesize a macro histological image from 3D histological slide deck, which compensates for the incomplete slides

Multi-step registration

Since mmSFE can capture reflectance and fluorescence images concurrently, there is no spatial registration error between these two images. The surface macro histological image was registered to the mmSFE reflectance image captured from the 2 mm brain slice, which contains more structural information for registration. Step 1, the boundary of the tissue was identified by leveling the images using Matlab function *graythresh*. Binary masks of tissue for both reflectance image and macro histological image were generated. Matlab function *imregtform* was used to register these two binary masks for affine transformation. The obtained affine transformation matrix T1 was applied to the fluorescence image captured from the same 2 mm slice using *imwarp*, Step 2, structural landmarks were selected in both macro histological image and mmSFE transformed reflectance image from Step 1 to serve as control points for a piecewise linear registration using Matlab function *cpselect* and a deformed transformation grid T2 was generated using Matlab function *fitgeotrans*. The process of cutting, fixation, and staining of the tissue to create histological slides introduces local deformations, which cannot be corrected using affine transformation. The piecewise linear registration divides images into triangular meshes, which are individually registered through linear transformations [87]. Small spacing of control points allows modeling of local deformation and can address local geometric distortions, while large spacing is used to model global deformations. Finally, the deformed grid T2 was applied to the transformed fluorescence image from step one. The entire pipeline is shown in Figure 3.5 below.

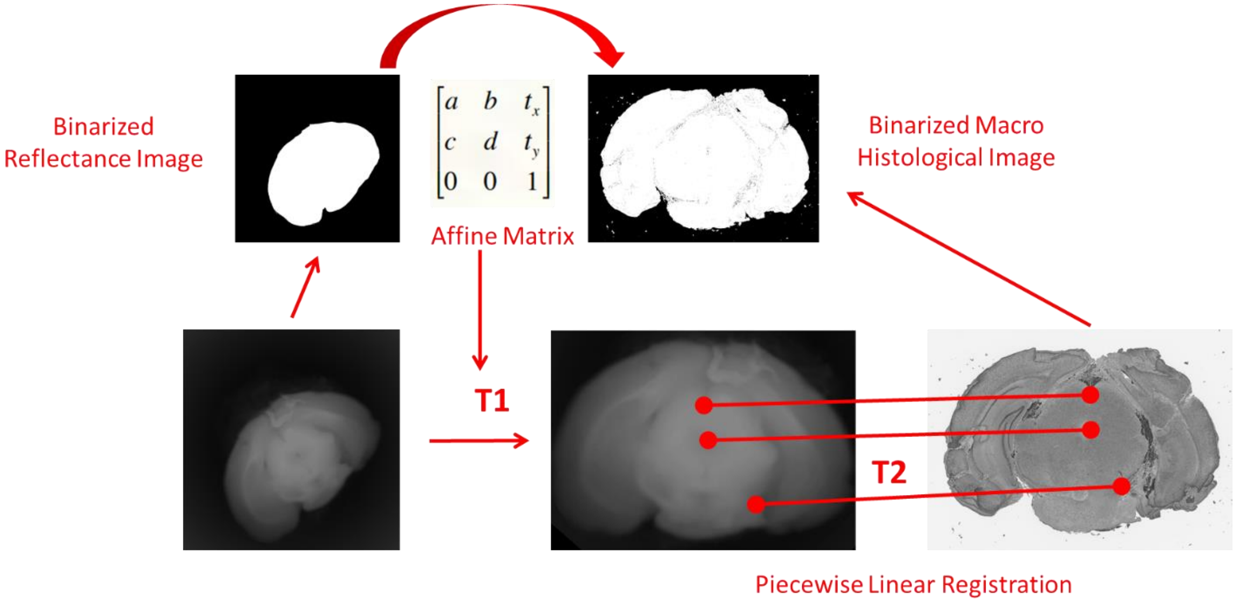


Figure 3.5 Pipeline for multi-step registration between surface macro histological image and mmSFE reflectance image: Affine transformation and piecewise linear transformation are applied to correct global transformations (rotation, translation, scaling) and local distortions, respectively. The obtained affine transformation matrix T1 and deformed grid T2 is applied to both reflectance and fluorescence image.

2D tumor map synthesis

A method based on optical attenuation model and AIP was applied to generate a projected 2D tumor map, which integrated tumor information from different tissue depths. For each digitized H&E stained image in a 3D histological slide deck, the tumor profile was outlined by an animal pathologist technician with extensive experience in mouse brain tumor. The outlined tumor profiles were extracted to form a 3D tumor map deck. Then optical attenuation model was used to simulate attenuated fluorescence emitted from deeper tumor before projecting the tumor maps to the outer surface. The details of these steps are described below.

The tumor regions were outlined from the 3D slide deck, and an optical attenuation model was applied to this 3D slide deck to compensate the scattering and attenuation in tissue. The effective attenuation coefficient can be calculated as follows:

$$\mu_{\text{eff}} = \sqrt{3\mu_a(\mu_a + \mu_s')}$$

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

Where $\mu_{\text{a}}(cm^{-1})$ is the absorption coefficient equal to 0.1 in biological tissue [31]; g is anisotropy factor equal to 0.9 [30] and $\mu_{\text{s}}(cm^{-1})$ is the scattering coefficient equal to 125 cm^{-1} estimated for cortical layers from 300 to 1500 μm in the rodent brain in the NIR range of 830 nm [74].

The attenuated intensity for each pixel can be calculated as:

$$I = I_0 * e^{-\mu_{eff}d} \tag{2}$$

Where I_0 is the initial intensity of the emitted fluorescence, and d is the distance travelled through the tissue. In this paper, we used $\mu_{\text{eff}}(cm^{-1})$ equal to 1.94 cm^{-1} for shallow cortical layers and 3.34 cm^{-1} for depths greater than 1500 μm, measured by Abdo et al in the NIR range of 830 nm [74].

After the attenuation model was applied to 3D tumor map deck, AIP was used to project attenuated tumor maps to the outer surface, shown in Figure 3.6. The synthesized 2D tumor map was used as ground truth and compared with the registered fluorescence image captured from the 2 mm slice.

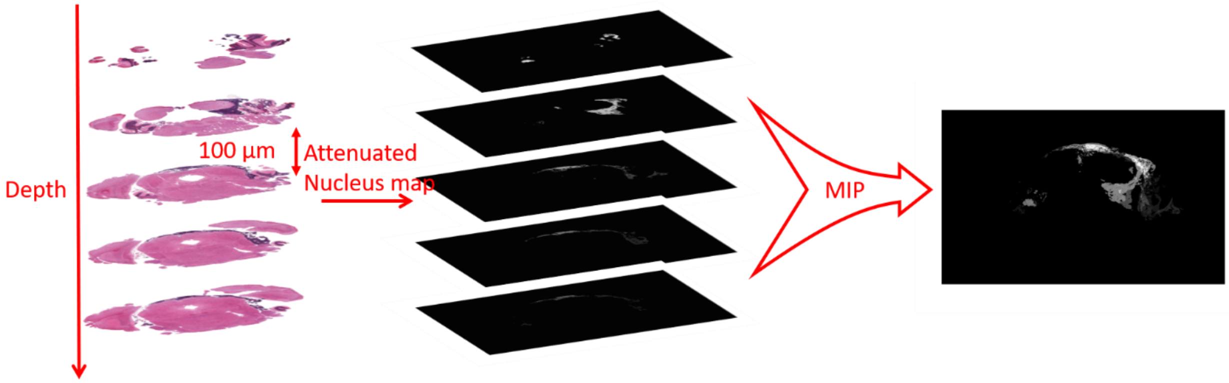


Figure 3.6 2D tumor map synthesis using optical attenuation model and AIP: An optical attenuation model is applied to the tumor regions extracted from 3D H&E slide deck. AIP is used to integrate attenuated tumor maps from different tissue depths as a 2D tumor map, which is later used as ground truth and compared with fluorescence.

3.2.6 *Performance metric and statistical analysis*

To evaluate the proposed pipeline in correlating fluorescence images with histological images and to evaluate BLZ-100 in tumor margin identification, brain slices with tumor and fluorescence labeling were included in the performance analysis. Slices with non-specific binding fluorescence were excluded. Figure 3.7 shows an example of non-specific binding, where BLZ-100 was accumulated in the choroid plexus, which is identifiable from other brain tissues during surgery. For each included slice, the synthesized 2D tumor map was used as ground truth. An additional region-of-interest representing the tumor margin was created from the 300 µm peripheral aspect of the 2D tumor map using Matlab function *imerode*. Five thousand pixels, accounted for 4.5 mm^2 in area, were randomly selected from the tumor core, tumor margin and background regions, enabling objective comparisons of the fluorescence. Target to background (T/B) ratio was then calculated in the tumor core and margin for each slice. Fluorescence intensity was used for fluorescence image thresholding to delineate tumor boundary. A pixel-wise comparison between 2D tumor map and thresholded fluorescence image was conducted. The optimal threshold was determined using Youden’s index of the receiver operating characteristic (ROC) curve [75], which optimizes the sum of sensitivity and specificity. Sensitivity, specificity, and Dice similarity coefficient (DSC), which is a popular metric for evaluating image segmentation and registration [85], were calculated accordingly. Area under the curve (AUC) of ROC curve was performed to determine the accuracy to differentiate between normal tissue and tumor, which evaluates the proposed methods independently of the threshold. To evaluate the performance of using synthesized 2D tumor map as ground truth, the average sensitivity, specificity, DSC and AUC were calculated and compared to those of using selected best slide (first complete slide in the 3D slide deck) and randomly selected slide from the 3D slide deck as ground truth. Group averages and variances were analyzed using paired samples t-tests and F-test to assess differences in the performance using stacked slides, selected best slide and randomly selected slide as ground truth, in terms of the accuracy in differentiating between tumor and normal brain.

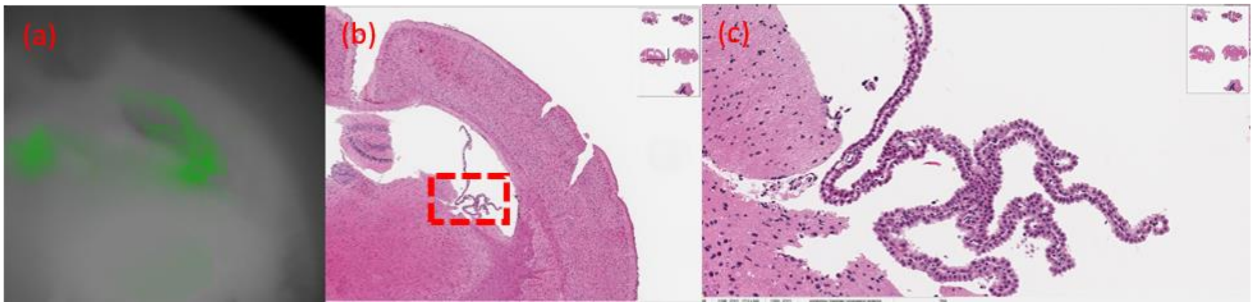


Figure 3.7 Example of BLZ-100 non-specific binding in the choroid plexus: (a) BLZ-100 binds to choroid plexus shown in fluorescence image (b-c) Corresponding histological image of the choroid plexus

3.3 RESULTS

3.3.1 *Distance compensation performance*

The proposed method, to compensate for inconsistent distance between the scope and tissue, was tested using a uniform gelatin phantom. mmSFE scope tip was placed perpendicular to the phantom at 1 cm away. Fluorescence and reflectance images were captured and combined as a ratio image as shown in Figure 3.8 (a – c). A 20 pixel-wide line profile was generated for each image at the same coordinates. The average intensity of the 20 pixels was plotted along the profile, which shows large variations in fluorescence (blue) and reflectance (red). The fluorescence and reflectance intensity at the image center are much greater than that at the periphery. The combination of reflectance and fluorescence as a ratio image (yellow) reduces these variations along the profile (Figure 3.8 (c-d)).

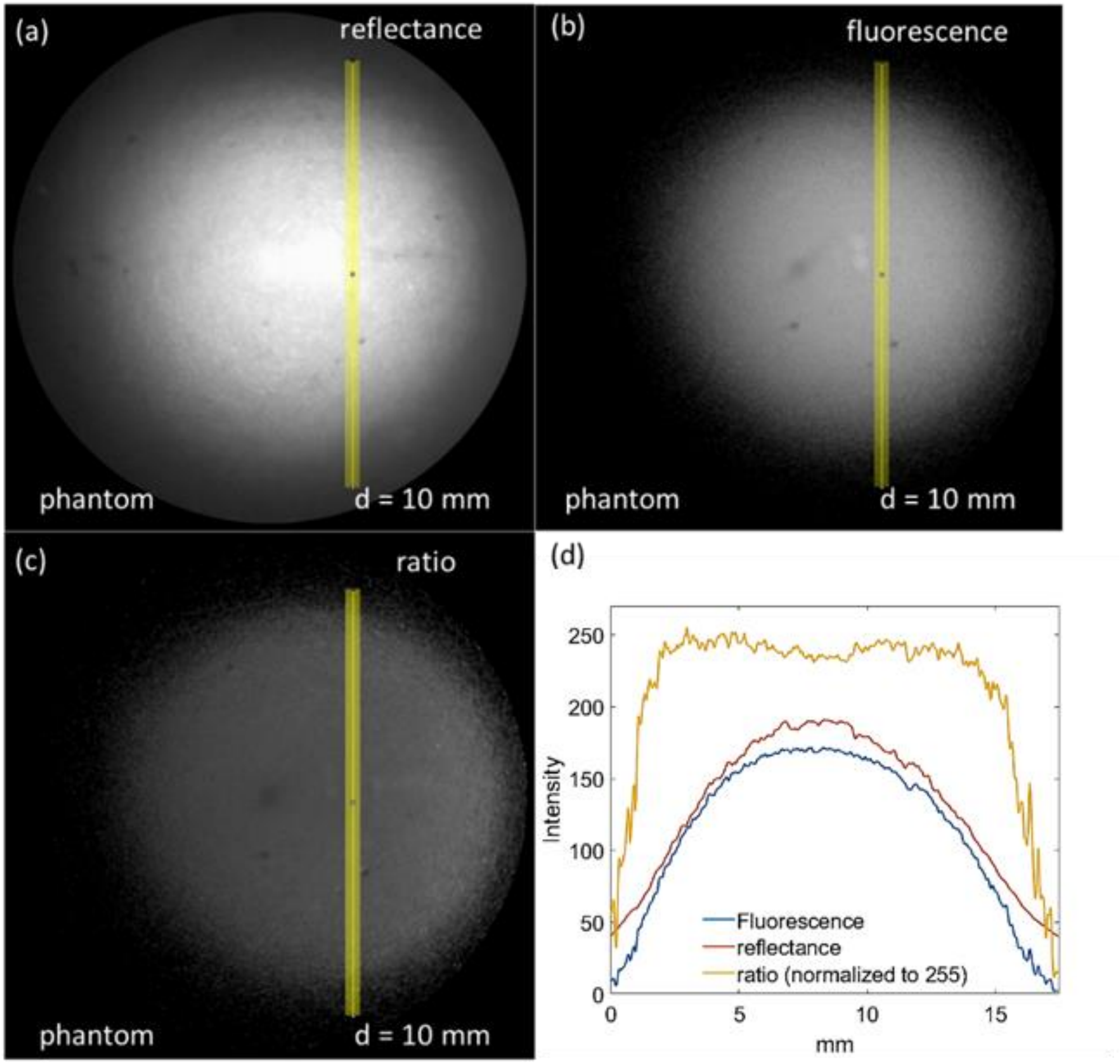


Figure 3.8 Proposed distance compensation method tested using a uniform gelatin phantom mixed with ICG (1 μM). The (a) reflectance (b) fluorescence images collected using mmSFE from the gelatin phantom at 10 mm away. (c) The combined ratio image using reflectance and fluorescence image. (d) The average intensity of 20 pixel plotted through line profiles in (a) - (c) shows large variations in fluorescence (blue) and reflectance (red) along the profile. The combined ratio image (yellow) reduces these variations.

3.3.2 *Multi-step image registration*

The surface macro histological image was created from 3D H&E slide deck and registered to calibrated mmSFE reflectance image from 2 mm slice. Affine transformation was applied to correct global deformations, such as scale difference, orientation difference, while piecewise linear registration was applied to correct local deformations. Figure 3.9 shows examples of registration results: original mmSFE reflectance image before registration (a, f, k); mmSFE green fluorescence image overlaid on reflectance image after registration (b, g, l); thresholded mmSFE fluorescence image overlaid on reflectance image after registration (c, h, m); synthesized 2D green tumor map overlaid on grayscale macro histological image (d, i, n); manually delineated tumor map overlaid on selected best slide (e, j, o). Alpha blending was applied to overlay fluorescence image over reflectance, which maintains the structural information present in the reflectance while adding the molecular information of the fluorescence distribution. As can be seen in Figure 3.9 (n), the synthesized 2D tumor map projected deeper tumor directly onto the surface. Tumor map from selected slide may underestimate the tumor information as seen in Figure 3.9 (e, j, o). Fluorescence hotspot colocalized well with the tumor regions on the synthesized 2D tumor map. However, the spatial distribution of BLZ-100 is not uniform. BLZ-100 showed strong accumulation in tumor core, while the margin appears more diffuse, as seen in Figure 3.9 (b, g, l).

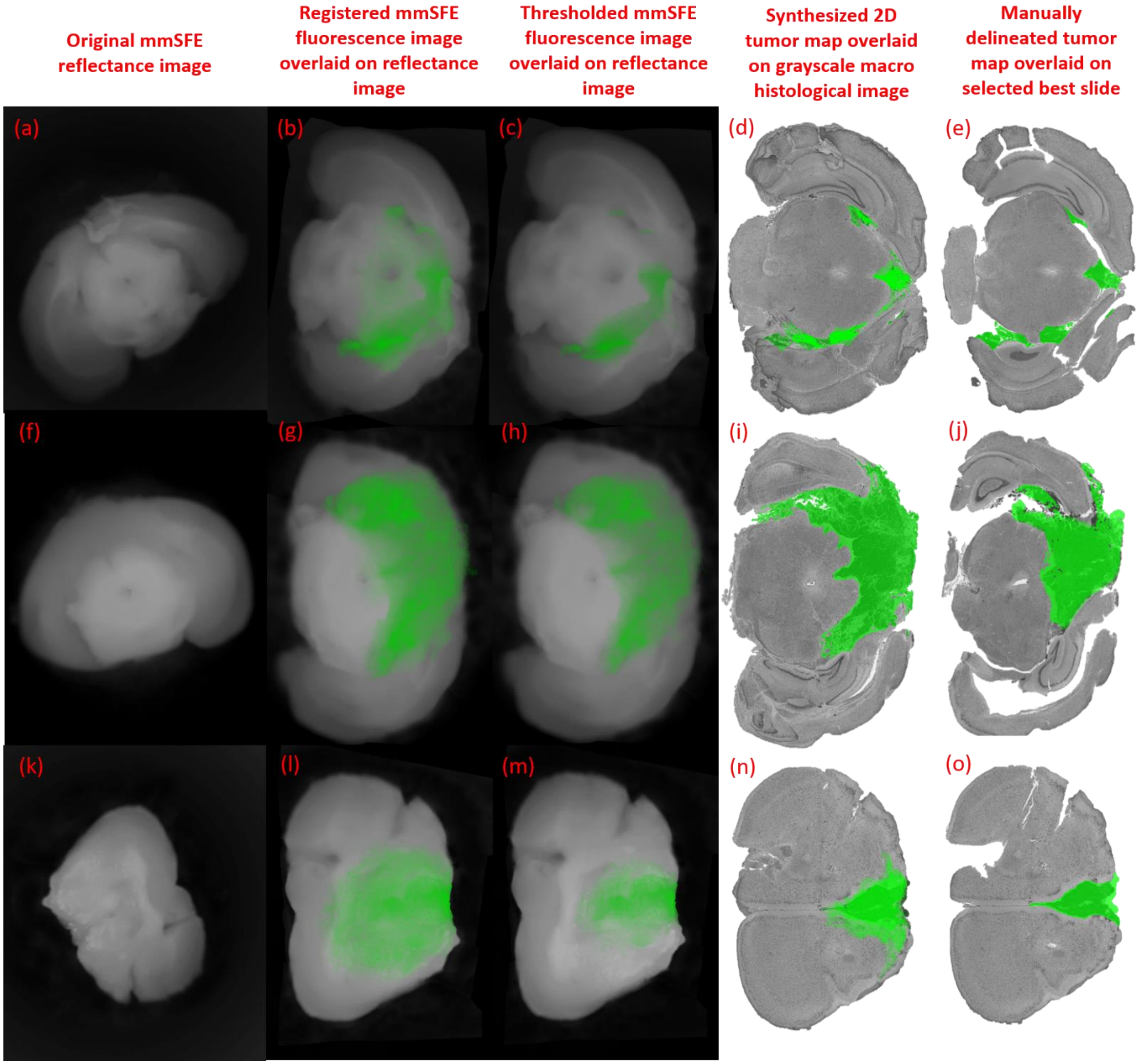


Figure 3.9 Examples of multistep registration: (a, f, k) original mmSFE reflectance image; (b, g, l) mmSFE fluorescence image overlaid on reflectance image after registration; (c, h, m) thresholded mmSFE fluorescence image overlaid on reflectance image after registration; (d, i, n) synthesized 2D tumor map using overlaid on macro histological image; (e, j, o) manually delineated tumor map overlaid on selected best slide.

3.3.3 *Overall framework performance and statistical analysis*

Fifty brain slices were used for mmSFE fluorescence imaging. Four slices became incomplete during the histological process. The confusion matrix of the remaining 46 slices is shown in table 3.1. The number of slices with tumor that were successfully labeled using BLZ-100 was 29. These true positive cases were used to evaluate this fluorescence imaging agent in tumor margin identification. For true negatives, six slides had no tumor present and displayed no fluorescence label. Eight slides with non-specific BLZ-100 binding in the choroid plexus were excluded in performance analysis. Three false negative cases were found, where little or no fluorescence was observed in slices with tumor.

Table 3.1 Confusion matrix of BLZ-100 performance in labeling medulloblastoma tumor model (Med-114FH)

	Condition positive (Slices w/ tumor)	**Condition negative (Slices w/o tumor)**
Test outcome positive (w/ fluorescence)	29	8
Test outcome negative (w/o fluorescence)	3	6

For the 29 included true positive cases, T/B ratios for the tumor core and margin were calculated using 5000 different points ($4.5\ mm^2$) randomly selected from tumor core, margin and normal regions. The mean T/B ratios (mean ± standard deviation) for tumor core and margin are 8.64 ± 5.76 and 4.82 ± 2.79, respectively, with p-value 1.86e-06 determined by paired t test. Fluorescence intensity was used to threshold tumor boundary and ROC curve was generated accordingly. ROC curve analysis showed marked difference in the performance between using integrated stacked slides as ground truth, using selected best single slide as ground truth and randomly selected single slide as ground truth, in terms of differentiating between tumor and healthy tissue. Mean sensitivity, specificity, DSC and AUC of ROC curve were investigated to compare the methods to generate tumor map. The synthesized 2D tumor map performed better than the selected best slide and randomly selected slide, as shown in table 3.2 and Figure 3.10. The average DSC for synthesized 2D tumor map was significantly greater than that for selected best slide (0.582 vs. 0.398 with p-value 2.36e-09) and that for randomly selected slide (0.582 vs. 0.396 with p-value 2.28e-07). The average AUC for synthesized tumor map was greater than that for selected best slide (88% vs. 85.5%), but there was no statistically significant improvement. The average AUC for synthesized tumor map was significantly greater than that for randomly selected slide (88% vs. 84.1% with p-value 2.18e-03). The variance of the AUC using synthesized tumor map as ground truth is significantly less than that using selected slide as ground truth (0.057 vs. 0.087, p-value 0.0145) and that using randomly selected slide as ground truth (0.057 vs. 0.085, p-value 0.0212).

Table 3.2 Mean sensitivity, specificity, DSC and AUC of ROC curve in differentiated between tumor and healthy tissue compared between synthesized 2D tumor map and selected slide

	Sensitivity	**Specificity**	**DSC**	**AUC**
“All” slides (synthesized 2D tumor map)	0.819 (0.797 - 0.842)	0.821 (0.795 - 0.846)	0.582 (0.510 - 0.654)	0.880 (0.859 - 0.901)
Selected best slide	0.817 (0.781 - 0.852)	0.800 (0.773 - 0.831)	0.398 (0.322 - 0.474)	0.855 (0.823 - 0.887)
Randomly selected slide	0.817 (0.768 - 0.866)	0.784 (0.734 - 0.834)	0.396 (0.322 - 0.470)	0.841 (0.810 - 0.872)

NOTE: The 95% CI of the mean is indicated in the parentheses

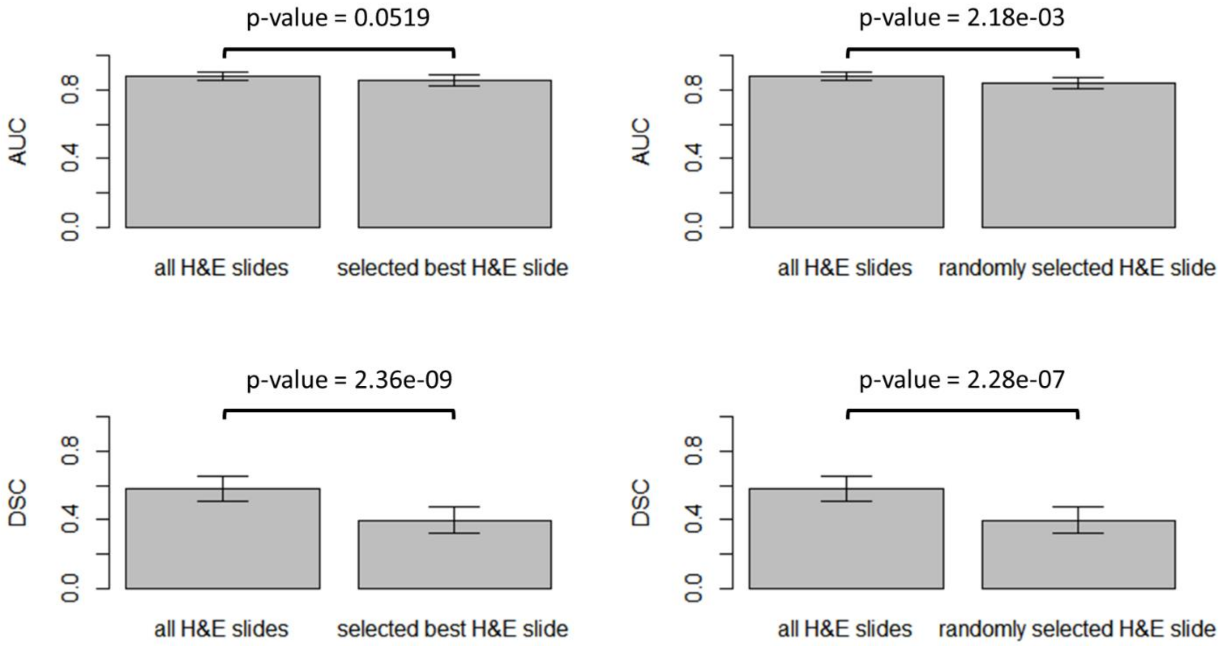


Figure 3.10 Mean AUC of ROC curve (Top row) and Mean DSC (Bottom row) in differentiated between tumor and healthy tissue compared between synthesized 2D tumor map, selected best slide and randomly selected slide. Significant differences between the groups is determined by paired t test. The 95% CI of the mean is indicated by the error bar.

3.4 DISCUSSION

In this chapter, a pipeline was developed to evaluate the performance of fluorescence imaging agents, in terms of tumor identification. This work was divided into two parts, first the fluorescence image was calibrated for accurate fluorescence quantification. As stated in [14], the performance of fluorescence imaging is affected by the variable experimental settings, including lesion depth and camera–tissue distance. To connect fluorescence imaging to clinical outcome and to ensure a reproducible performance, calibration of fluorescence imaging is required. To address this problem. A specular reflectance rejection algorithm was applied, and a background subtraction method was used to reject ambient light. Inconsistent distance and angle were then compensated by combining reflectance image and fluorescence as a ratio image. When testing in the gelatin phantom, the ratiometric method compensated nonuniform fluorescence. As can be seen on the intensity plot (Figure 3.8), fluorescence was much higher in the center of the image than at the periphery. The ratio image corrected for these differences over the width of the image and the uniform fluorescence (80% -100% of the maximum fluorescence) was extended from 50% to 80%. Secondly, a multi-step registration framework was established to correlate fluorescence images to stacked histological images. To evaluate an imaging agent preclinically, a proper registration between the fluorescence and histological images is needed. However, current registration methods compare to single selected H&E slide [69, 70], which is subjective and may underestimate fluorescence from deeper lesions, especially for NIR fluorescence agents. In this paper, an optical attenuation model and AIP was applied to integrate multiple histological slides and project deeper tumor regions to the surface. One advantage of this proposed method is the use of stacked H&E slides from surface to deeper tissue, to generate a synthesized 2D tumor map, which provides an approximation of light in 'seeing' tumor from deeper tissue. Then global and local deformations were corrected using affine transformation and piecewise linear registration. Structural landmarks from the reflectance image were selected as control points for piecewise linear registration.

NIR agent, BLZ-100 was used to evaluate the proposed pipeline. The 2D synthesized tumor map was compared to the selected slide as ground truth. In this study, we selected the complete slide with the best quality and a random slide in the 3D slide deck as comparison. Sensitivity, specificity, DSC and AUC were calculated as evaluation metrics. The optimal fluorescence intensity threshold for tumor delineation was calculated based on Youden's index, which optimizes the sum of sensitivity and specificity. There is no significant difference in sensitivity and specificity when comparing the 3 methods. However, we obtained a significantly higher DSC when using the synthesized tumor map as ground truth. The reasons are illustrated as follows: (1) The selected slide only accounts for a single plane from the 3D tumor, which underestimates the tumor information. The underestimated tumor map results in a reduced number of true positive pixels and an increased number of negative pixels, as seen in Fig. 10 (e, j, o). In contrast, there are more true positive pixels and fewer negative pixels when using synthesized 2D tumor map as ground truth. (2) The imbalanced pixel number between normal tissue and tumor aggravates the discrepancies in the ratio between the number of negative pixels and the number of true positive pixels (synthesized tumor map vs. selected best slide: 7.42 vs. 18.57, with p-value 2.55e-05 determined by paired t test). The significant discrepancies in the ratio between the number of negative pixels and the number of true positive pixels results in a significant difference in DSC, shown in the equation below.

$$DSC = \frac{2}{1 + (1 - specificity) * \frac{N}{TP} + \frac{1}{sensitivity}}$$

Where N is the number of negative pixels and TP is the number of true positive pixels. This imbalance between the number of true positive pixels and the number of negative pixels diminishes the robustness of using Youden’s index to select optimal threshold. AUC is a more reliable metric to evaluate the proposed methods, which is independent of the selected threshold. The higher average DSC and AUC suggests a better spatial overlap between BLZ-100 fluorescence and synthesized tumor map from H&E slides. Especially, the proposed method outperforms the randomly selected single plane, which is usually obtained from regular pathological processes, with significantly higher DSC and AUC. When compared to using selected best slide and randomly selected slide as ground truth, the significantly smaller variance of the AUC using synthesized 2D tumor map as ground truth indicates that this method is less biased, more stable and more reproducible. Overall, the 2D synthesized tumor map outperformed the selected best slide and randomly selected slide when comparing with the NIR fluorescence image, which is also a 2D mapping of a 3D tumor.

The experiment is conducted on ex-vivo tissue as preclinical evaluation. The performance of this pipeline needs to be investigated for in-vivo intraoperative imaging, which can be challenged by the complex tissue properties and the variable experimental arrangement. For example, we fixed the mmSFE position with one perspective for tissue imaging in this study. During freehand surgical applications, mmSFE position will be adjusted for multi-perspective views in real-time video. The change of scope-tissue angle may diminish the importance of the proposed distance compensation method. One assumption of the proposed method is that the fluorescence and reflectance light travel along identical paths to the scope, which is valid for neoplastic lesions topically labeled from the epithelial layer [35]. However, systemic labeling of lesions using NIR fluorescence can emit from deeper tumors, which makes the light paths not identical to the scope. A tilted orientation of scope to tissue surface may magnify this error. On the other hand, the multi-perspective views can be used for elimination of specular reflections without loss of information. Current specular removal method uses surrounding pixels to replace high intensity pixels from specular reflection, which only corrects high intensity, but neglects the original structural features beneath the specular reflection. Future work is needed to design new specular reflection rejection algorithm, which corrects specular reflection using the same region from a different perspective view.

One limitation of this study is that only the 29 slices with tumor and fluorescence labeling were included in the performance analysis, which limited a comprehensive evaluation of BLZ-100 in tumor identification. However, we are evaluating the proposed pipeline in correlating fluorescence images to stacked histological images instead of testing the clinical effectiveness of BLZ-100. In addition, 5 H&E slices were integrated to generate a synthesized 2D tumor map, which accounted for only 0.5 mm in depth. Depending on the type of tissue, the penetration depth of light is less than 1 mm at 400 nm and increases to about 5mm at 700 to 900 nm using skin as an example [77]. The NIR fluorescence detection depth of the mmSFE was measured using ICG solution from 100 μM to 100 nM diluted in Dimethyl sulfoxide (DMSO). ICG solutions from 100 μM to 100 nM in microhematocrit capillary tubes (Fisher Scientific, 1.1 mm ID) were buried in NSG mouse brain (0.5 mm, 1 mm, 2 mm, 3 mm from brain surface). mmSFE was used to capture the ICG fluorescence at 1 cm away from the brain surface. T/B ratio was calculated accordingly. The mmSFE detection depth is shown in Table 3.3. From the table, we can see that mmSFE can detect 100 nM ICG at a depth of 0.5 mm and 1 μM ICG at a depth of 1 mm in mouse brain tissue. Since the estimated BLZ-100 dye concentration accumulated in the tumor is up to the equivalent of 1 μM, applying 5 layers of H&E slices (total depth of 0.5 mm) to synthesize the 2D tumor map is underestimated. To better represent the information from deeper lesions detected by the NIR probes, more histological slices, 1 mm in total depth, should be integrated as a synthesized 2D tumor map and compared with the mmSFE NIR fluorescence image.

Table 3.3 mmSFE detection depth measured in T/B ratio

	100 µM	**10 µM**	**1 µM**	**100 nM**
0.5 mm	23.83	14.71	9.38	4.95
1 mm	13.78	12.38	4.34	---
2 mm	9.34	4.20	---	---
3 mm	4.89	---	---	---

NOTE: --- indicates that fluorescence signal cannot be differentiated from noise

Moreover, the optical attenuation model is limited by the assumption of straight-line fluorescence attenuation (orthogonal to tissue surface) which is incorrect for highly scattering brain tissue. More complex models, such as Monte Carlo light scattering simulation, can be used to better simulate fluorescence scattering in tissue [78]. In contrast, maximal intensity projection (MIP) can be regarded as a simplified model, which has similar results compared to the optical attenuation model. However, the 2D tumor map synthesized using MIP does not include tumor depth information. Instead, the optical attenuation model and Monte Carlo light scattering simulate the fluorescence scattering in the tissue, which can provide more realistic tumor maps including tumor depth information. Last but not least, we used single NIR imaging agent for one type tumor detection in one mouse model. The efficiency of the proposed pipeline in calibrating different imaging agents to histology for different tumor types, needs to be further investigated. In our future work, we aim to automate the whole pipeline, thus providing reliable tool to evaluate fluorescent probes in delineating margins. The calibration techniques can be realized in real-time and implemented using mmSFE for standardized in vivo fluorescence imaging, along with pre- and possibly post-procedure calibrations. The multi-step registrations can be automated using laser-induced fiducial landmarks [79] and algorithm-selected landmarks [80].

When BLZ-100 NIR fluorescence is compared to three other fluorescence agents in Elliott et al., (2016) [69], in general their performance in detecting glioma tumor area is comparable. In terms of mean accuracy determined by ROC curve analysis for wild-type tumor, BLZ-100 performed slightly better than the single fluorescence agents (BLZ-100: 88% vs. IRDye680: 77% vs. PpIX: 87% vs. ABY029: 69%), but slightly worse than the combination of these agents (combination: 90%) [69]. In terms of enhanced fluorescence contrast, BLZ-100 accumulates in both tumor cores and tumor margins with mean T/B ratio 8.64 and 4.82, respectively. The tumor core showed 1.8-fold higher TB ratio values compared with tumor margin (p-value 1.861e-06 determined by paired t test). The margin fluorescence appears more diffuse, which could be scattering fluorescence from deeper tumor lesions into non-tumor regions closer to the surface. While PPIX shows weak contrast in tumor margins and appears to contain more high spatial frequency features, sharper edges and patterns; IRDye680RD enhances the necrotic regions provided by the EPR effect; and ABY-029 shows enhanced fluorescence in both tumor core and margin, but limits to EGFR overexpressed tumor type [69]. As can be seen, these imaging agents can provide complementary properties (metabolism, perfusion, and EGFR expression) and may be combined with BLZ-100 in the future to provide more accurate tumor contrast for guiding surgery [86]. For example, in addition to the systemically applied BLZ-100, the QRH-882260 + Cy5 peptide [83] can be topically applied intraoperatively to the residual tumor during surgery to provide better visualization of tumor margins. The adoption of multiagent strategy for more accurate tumor delineation can be achieved by using mmSFE, which can detect multiple spectral bands of fluorescence concurrently. There is essentially no registration error across fluorescence channels, which makes multiagent imaging using mmSFE more powerful. Moreover, mmSFE can be modified for an additional OCT channel [81]. The combination of en face fluorescence imaging and 3D morphological imaging allows molecular and structural contrast simultaneously, which has been recently demonstrated using a xenograft mouse model of human colorectal cancer ex vivo [82].

Chapter 4. QUANTITATIVE MULTIPLEX FLUORESCENCE IMAGING FOR BARRETT’S NEOPLASIA DETECTION

In this chapter, I propose a method to quantitatively analyze multiplex fluorescence imaging for early Barrett’s neoplasia detection. Two fluorescence labeled peptides specific for EGFR and ErbB2 are applied to increase detection sensitivity.

4.1 BACKGROUND

In Chapter 2, a single fluorescence labeled peptide was used to detect esophageal neoplasia, and an automatic algorithm was developed to calculate T/B ratio, thus providing quantification for early cancer detection and potentially for guiding biopsy. Discussed in the last section of Chapter 2, the new imaging technology should have a recommended sensitivity of 90% or greater and maintain a specificity at least 80% for detecting HGD and EAC during the endoscopic surveillance of patients with BE as stated PIVI [40].

However, inter and intra-patient differences suggest that cancer detection targeting single biomarker is unlikely to provide the required specificity and sensitivity. Genetic changes that drive cancer transformation usually result in activation of multiple signaling pathways [88]. Detection of multiple targets is likely to be needed to adequately identify this disease process in the general patient population. Valdes et al. combined multiple biomarkers for in vivo quantification of high-grade glioma detection [89]. The precedent of using multiple biomarkers in detection is well-established in breast cancer. Three biomarkers used most in breast cancer assessment, are ErbB2, estrogen receptor (ER), and progesterone receptor (PR) [90]. EGFR and ErbB2 are transmembrane tyrosine kinase receptors that stimulate epithelial cell growth, proliferation, and differentiation

[91]. Overexpression of these targets reflects an increase in biological aggressiveness and higher risk for progression to Barrett's neoplasia. The genes for EGFR and ErbB2 have also been found to be high frequency amplified in surgically resected specimens of EAC [92-94]. In this study, by exploiting these two tumor-specific biomarkers, we expect to improve cancer detection sensitivity to 90% or greater, while maintain an 80% specificity.

Fluorescence labeled peptides QRH*-Cy5 and KSP*-IRDye800, targeting EGFR and Her2 respectively, are applied in this clinical trial. The broad spectrum of light over the visible and near-infrared (NIR) regime enables concurrent multiple target detection. mmSFE is applied to collect co-registered wide-field images of fluorescence and reflectance in separate channels. Quantitative analysis based on T/B ratios from high contrast regions is evaluated using support vector machine and ridged logistic regression with leave-one-out cross-validation.

4.2 MATERIAL AND METHODS

4.2.1 *Fluorescence-labeled peptides*

QRHKPRE was labeled with Cy5 via a GGGSK linker to prevent steric hindrance. This sequence was selected using phage display methods for specific binding to EGFR [83]. KSPNPRF (black) was labeled with IRDye800 via a GGGSC linker, hereafter KSP*-IRDye800. This sequence was identified using a structural model for specific binding to ErbB2 [95]. The fluorophores used were chosen to minimize spectral overlap between absorbance and emission. QRH*-Cy5 and KSP*-IRDye800 were synthesized using current good manufacturing practices (cGMP) methods. Peptide stability was monitored by visual appearance, purity (HPLC), and molecular weight over the duration of the study.

4.2.2 *mmSFE system for multiplexed imaging*

mmSFE was used to collect co-registered wide-field images of fluorescence and reflectance in esophagus. A custom fiber-coupled multi-laser system (SpectraTEC, Blue Sky Research) provides illumination at ex = 638 and 785 nm concurrently through a single mode fiber (SMF) to excite QRH*-Cy5 and KSP*-IRDye800, respectively. Large core (250 μm diameter) multi-mode fibers (MMF) with high numerical aperture (NA = 0.63, PJ series, Toray Industries Inc) are arranged around the instrument periphery to maximize light collection. In this study, a mmSFE with 2.1mm overall shaft diameter was used. The optics used to separate and collect fluorescence from each peptide was contained within the base station (VerAvanti). A notch filter (NF, #NF03-785E-25, Semrock) was used to attenuate the excitation beam at $\lambda = 785$ nm. Reflectance at $\lambda = 647$ nm was deflected by a dichroic beam splitter (BS_1, #FF652-Di02, Semrock) to a multi-pixel photon counter ($MPPC_1$, SiPM, Hamamatsu) for detection. Fluorescence from QRH*-Cy5 was deflected by a second beam splitter (BS_2, #Di02-R785, Semrock) centered at $\lambda = 785$ nm through a bandpass filter (BPF, FF01-697/58, Semrock) that transmits between 668-726 nm for detection by $MPPC_2$. Fluorescence from KSP*-IRDye800 passes through BS_2 and long pass filter (LPF, #BLP01-785R, Semrock) with cutoff at $\lambda = 805$ nm and is detected by $MPPC_3$. The detection scheme is shown in Figure 4.1 below.

4.2.3 *In vivo imaging*

Consecutive patients referred to Michigan Medicine for either evaluation or therapy of Barrett's neoplasia (HGD/EAC) were recruited. Inclusion criteria included: 1) history of BE/HGD/EAC, 2) medically able to tolerate upper endoscopy, and 3) age over 18. Exclusion criteria included: 1) history of esophagectomy and 2) on active chemotherapy or radiation protocol. The multiplexed imaging study was approved by the University of Michigan IRB and was registered online at ClinicalTrials.gov (NCT03589443). QRH*-Cy5 and KSP*-IRDye800 were reconstituted in 5 mL of normal saline separately, and topically administered into the distal esophagus using a standard spray catheter (PW-5V-1, Olympus Medical Systems Corp) passed through the instrument channel of the endoscope. Separate channels were used to record fluorescence images from QRH*-Cy5 and KSP*-IRDye800 and from reflectance at λ = 638 nm. The conventional white light illumination was shut off during collection of fluorescence images to prevent interference. Tissue specimens were collected with either biopsy or EMR from suspicious regions. The most advanced histological classification for each subject is recorded.

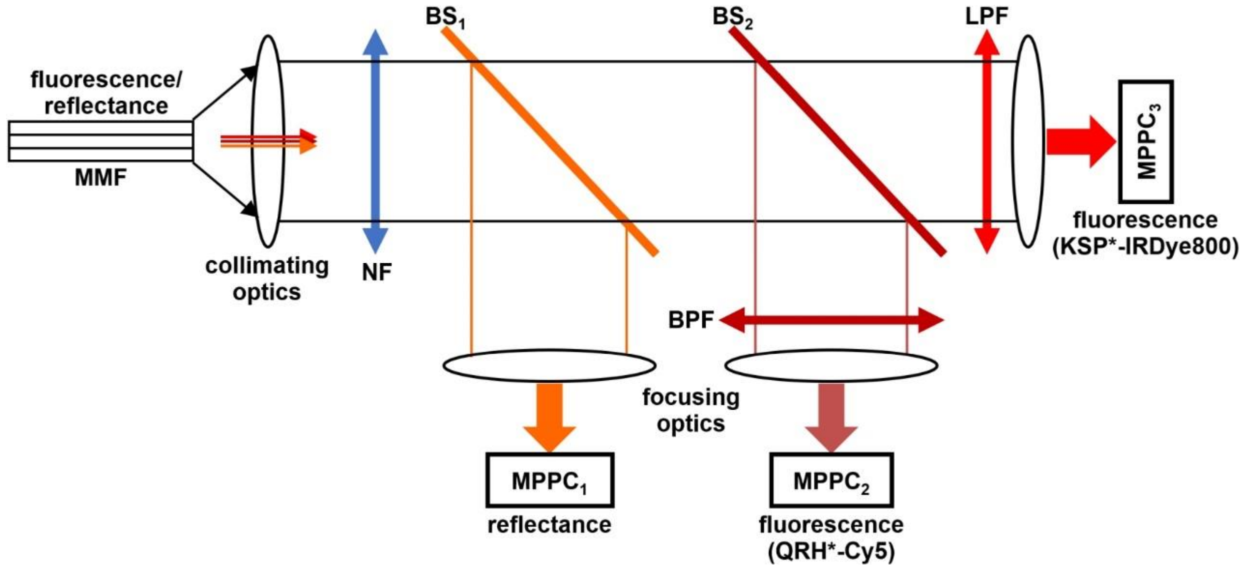


Figure 4.1 Detection scheme. Fluorescence and reflectance are collected and delivered by multi-mode fibers (MMF). A notch filter (NF) attenuates the excitation beam at λ = 785 nm. Reflectance is deflected by a beam splitter BS1 with center wavelength at λ = 647 nm and detected by a multi-pixel photon counter (MPPC1). Fluorescence from QRH*-Cy5 is deflected by a dichroic beam splitter (BS2) centered at λ = 785 nm. A bandpass filter (BPF) transmits fluorescence between 668-726 nm for detection by MPPC2. Fluorescence from KSP*-IRDye800 passes through BS2 and a long pass filter (LPF) with cutoff at λ = 805 nm for detection by MPPC3.

4.2.4 *Image analysis*

Video streams with >30 consecutive images (1 sec) having negligible artifacts, including debris, bubbles, pooling, and conventional white light, were extracted from the original recordings. Single frames from the target regions with high contrast and minimum motion blur and distortion were selected for intensity quantification. The frames were separated into two fluorescence channels and a co-registered reflectance image. Target/background (T/B) ratios were calculated for each fluorescence image as follows. A 5×5 Gaussian filter was used first to smooth the image. Fluorescence regions with high intensity were segmented using the Chan-Vese algorithm [96], where an active contour model was evolved iteratively to minimize the energy function, defined by the weighted sum of the 1) boundary length, 2) area, and variance of intensity 3) inside and 4) outside of the segmented region. The details of the algorithm can be found in Chapter 2.

After image segmentation, a series of standard morphology operations, including erosion and dilation, were implemented using OpenCV library [97], Erosion was performed to smooth the target region by removing isolated noise <10 pixels. Dilation was done to enlarge the boundaries of the target region and generate an adjacent 30 pixel-wide background [35]. T/B ratios were calculated from the mean intensities each region. The T/B ratios were used to classify each patient as being either positive (HGD/EAC) or negative (SQ/BE/LGD) based on the highest grade of pathology found. Different methods were investigated for classification, including the 1,2) T/B ratios from each fluorescence channel separately, 3) maximum T/B ratio, 4) logistic regression with L2 regularization which prevents overfitting of the model, and 5) support vector machine (SVM), which constructs a hyperplane for classification. The overall procedure for image analysis is illustrated in Figure 4.2.

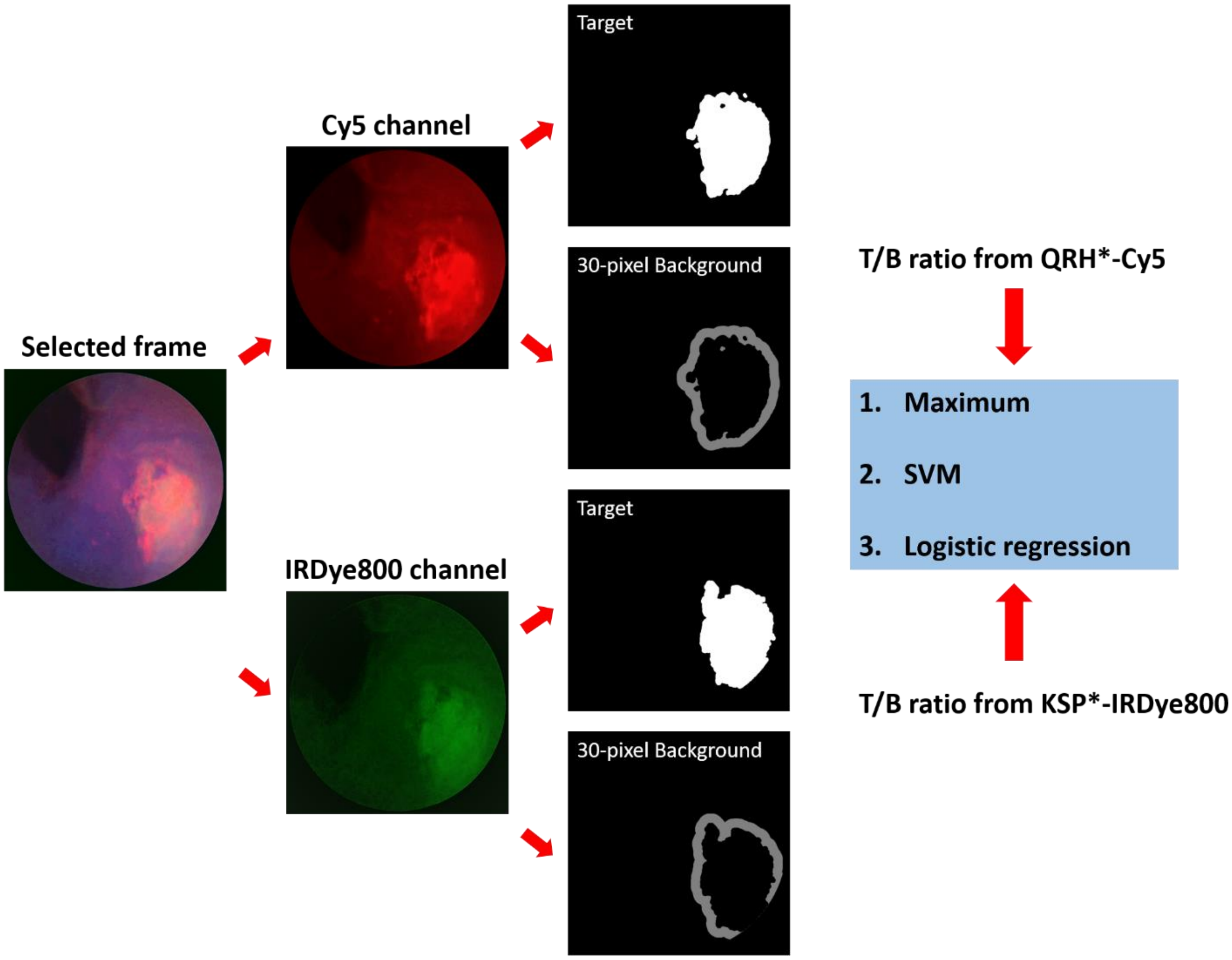


Figure 4.2 Analysis pipeline. Single frames from the target regions with high contrast and minimum motion blur and distortion were selected for intensity quantification. The frames were separated into two fluorescence channels, Cy5 and IRDye800. Target/background (T/B) ratios were calculated for each fluorescence image as the mean intensity in the target divided by the mean intensity in the 30-pixel wide surrounding background. The calculated T/B ratios were combined for classification using different methods.

A leave-one-out cross-validation (LOOCV) was performed to identify the best method [98]. For each round, the T/B ratios from n = 21 subjects were used as a training set to determine the discriminant function used to assign a score to the left-out subject. For methods 1-3, the maximum Youden index of the receiver operating characteristic (ROC) curve was used to select the optimal cutoff threshold for HGD/EAC, which maximizes the vertical distance between the ROC curve and the diagonal line [99]. Normally, Youden index = sensitivity + specificity – 1.

For logistic regression and SVM, a nested cross validation was performed to select the hyperparameters for each model, including C-values (inverse of L2 regularization strength) for logistic regression, γ-values (kernel coefficient), and C-values for SVM with a radial basis function (RBF) kernel. A repeated stratified 9-fold cross-validation in the inner loop was used to select the hyperparameters that provided the maximal classification accuracy for the 21 samples. The model was refit with these optimal values using n = 21 data points, and then the result left out in the outer loop was evaluated. The average classification accuracy of each method from LOOCV was compared. The sensitivity, specificity, and area-under-the-curve (AUC) for the ROC curve using LOOCV was computed and compared for logistic regression and SVM. The optimal model was then retrained using all data. The performance for the multiplexed detection using the trained optimal model was then compared with that for either QRH*-Cy5 or KSP*-IRDye800 alone, using average ROC curve from 1000 bootstrap samples. The analysis was implemented using Python using scikit-learn library [100].

To evaluate the reproducibility and possible bias in selecting a single high-quality frame as a representative for T/B calculation, multiple individual frames were extracted from representative video streams collected in vivo from the esophagus for expression of **A**) EGFR and **B**) ErbB2. The mean and standard deviation of replicate estimated T/B ratios are calculated and compared the T/B ratio of the representative frame.

4.3 RESULTS

4.3.1 *In vivo imaging*

IND #127,224 and #135,250 were approved by the FDA for use of QRH*-Cy5 and KSP*-IRDye800, respectively, in humans for research purposes. Multiplexed images were collected from a total of n = 22 subjects. QRH*-Cy5 and KSP*-IRDye800 were reconstituted in 5 mL of normal saline, and topically administered sequentially via a standard spray catheter. After 5 minutes for incubation, the unbound peptides were washed away with tap water using an irrigator. Residual fluid was removed by suction. The mmSFE was passed forward through the working channel of the endoscope into the distal esophagus. Co-registered fluorescence and reflectance images were collected. Representative white light images are shown for squamous (SQ) and BE, Figure 4.3 A,B. The location of BE was identified by a salmon pink appearance above the gastric folds with a minimum length of 1 cm. Following peptide administration, the background appeared minimal and relatively uniform. Fluorescence images were collected in each channel and were displayed separately. Co-registered reflectance provided anatomic landmarks on the merged images to facilitate interpretation. A representative set of in vivo images for HGD and EAC is shown, Figure 4.3 C,D. Increased fluorescence intensities can be seen from regions of HGD and EAC as confirmed by pathology.

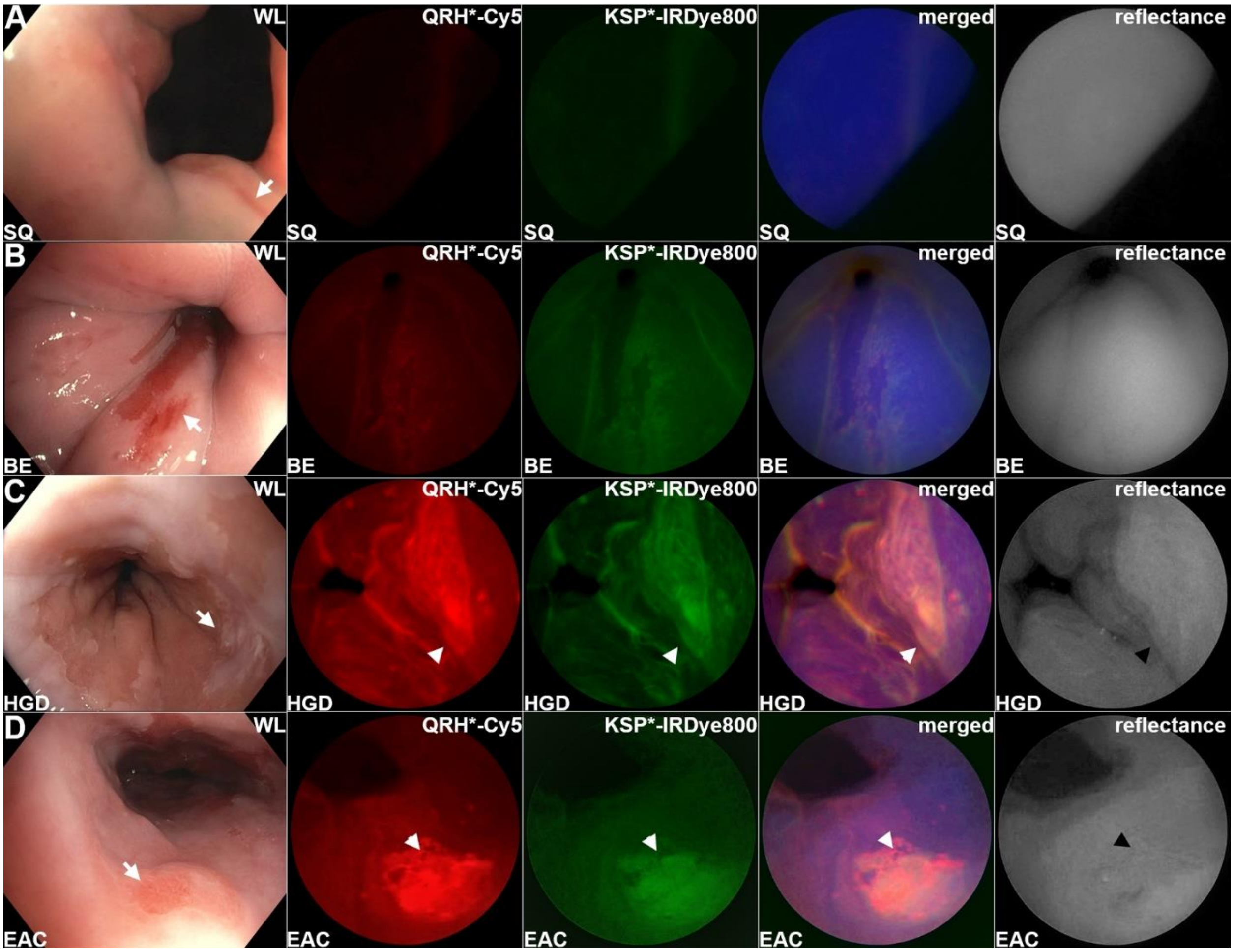


Figure 4.3 Barrett's esophagus. In vivo images collected endoscopically are shown from patients with **A**) squamous (SQ), **B**) Barrett's esophagus (BE), **C**) high grade dysplasia (HGD), and **D**) esophageal adenocarcinoma (EAC). The presence of BE is identified by the salmon red patches (arrows) on the white light images. Fluorescence images are collected after topical administration of QRH*-Cy5 and KSP*-IRDye800. The merged images show high contrast regions-of-interest (ROI) where EGFR and ErbB2 (orange) are co-expressed. Co-registered reflectance images provide anatomical landmarks to interpret the location of the ROI's.

4.3.2 *Image analysis*

The T/B ratios using QRH*-Cy5 and KSP*-IRDye800 were measured from individual patients, and are shown in the scatter plot, Fig. 4A. For SQ (n = 2) or NDBE (n = 3), a mean (±SD) T/B ratio of 1.28±0.07 for QRH*-Cy5 and 1.33±0.15 for KSP*-IRDye800, respectively, was calculated. The T/B for (n = 4) LGD was 1.23±0.05 and 1.18±0.10, respectively. For HGD (n=7) and EAC (n = 6), a mean (±SD) T/B ratio of 1.61±0.21 and 1.68±0.24, respectively, was found. The results were classified per the highest grade of pathology using several methods with leave-one-out cross-validation (LOOCV), Table S7. Support vector machine (SVM) and logistic regression (LR) provided the highest classification accuracy of 91%. The imaging results revealed n = 12, 1, 8, and 1 true positives, false positives, true negatives, and false negatives, respectively, resulting in 92% sensitivity and 89% specificity. The decision boundaries using SVM and LR are shown, Figure 4.4 A,B. The ROC curves using these methods from LOOCV are displayed, Figure 4.4 C. A higher AUC was achieved with multiplexed detection versus either target alone from bootstrap, Figure 4.4 D. The T/B ratios from individual images extracted from video streams showed that use of representative high-quality images resulted in only a small bias that was similar for the two targets, Figure 4.5.

Table 4.1 In vivo imaging results

Methods	Accuracy	AUC
1) T/B ratio from QRH*-Cy5	86%	0.95
2) T/B ratio from KSP*-IRDye800	82%	0.94
3) Maximum T/B ratio	86%	0.96
4) Support vector machine (SVM)	91%	0.97
5) Logistic regression	91%	0.97

Notes: Classification accuracy for in vivo images versus highest grade of pathology is shown for different methods using leave-one-out cross-validation (LOOCV). Average AUC with 1000 bootstrap samples using the models trained on all of the data is shown. SVM and logistic regression models combine T/B ratios from QRH*-Cy5 and KSP*-IRDye800.

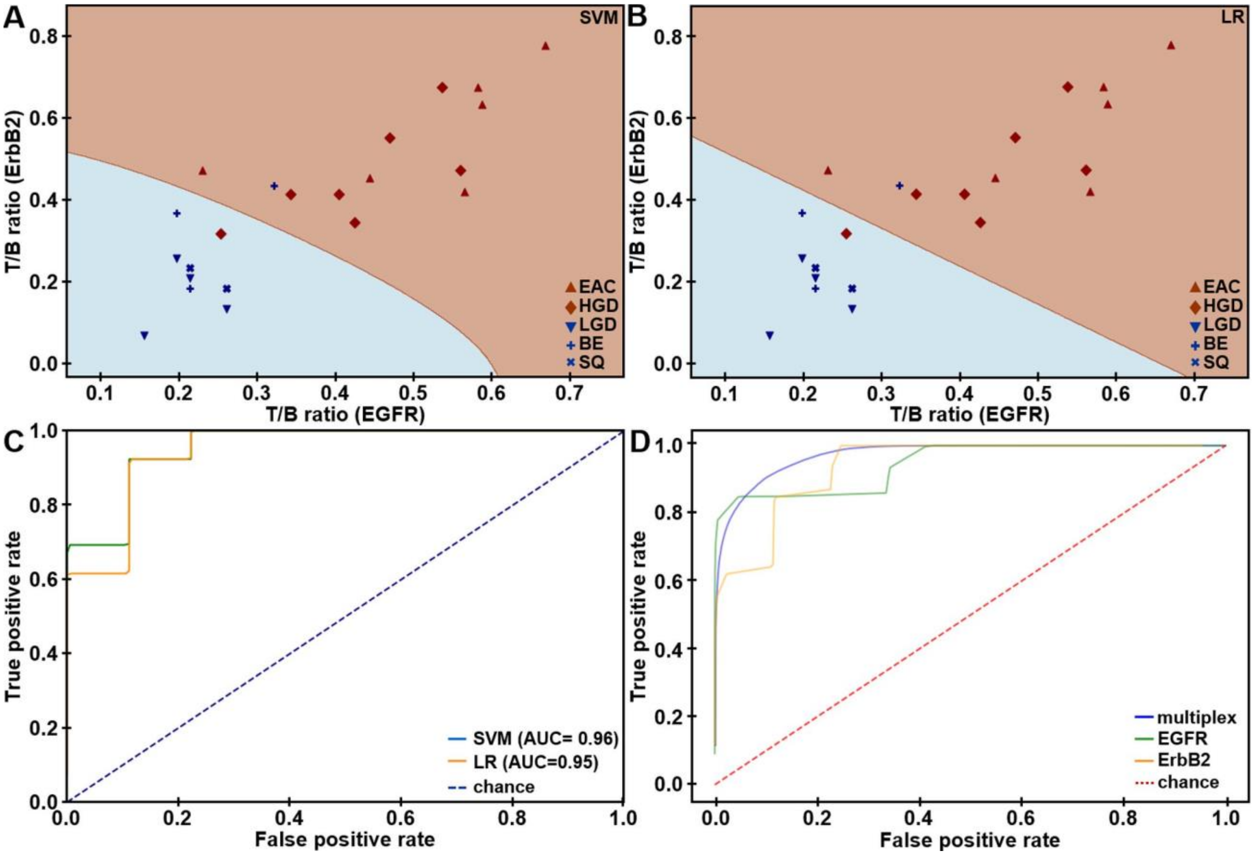


Figure 4.4 In vivo imaging performance. Scatter plot shows T/B ratios measured for EGFR and ErbB2 expression in the fluorescence images collected in vivo from the distal esophagus of n = 22 patients. Decision boundaries are shown for **A**) support vector machine (SVM) and **B)** logistic regression (LR) trained on all data. **C)** ROC curves for classifying HGD/EAC from SQ/NDBE/LGD are shown using SVM and LR algorithms with leave-one-out cross-validation (LOOCV). **D**) Average ROC curves from bootstrap using SVM (AUC = 0.97) and LR (AUC = 0.95) model trained on all data show that multiplexed detection provides better performance than detecting either EGFR or ErbB2 alone (AUC = 0.94).

.

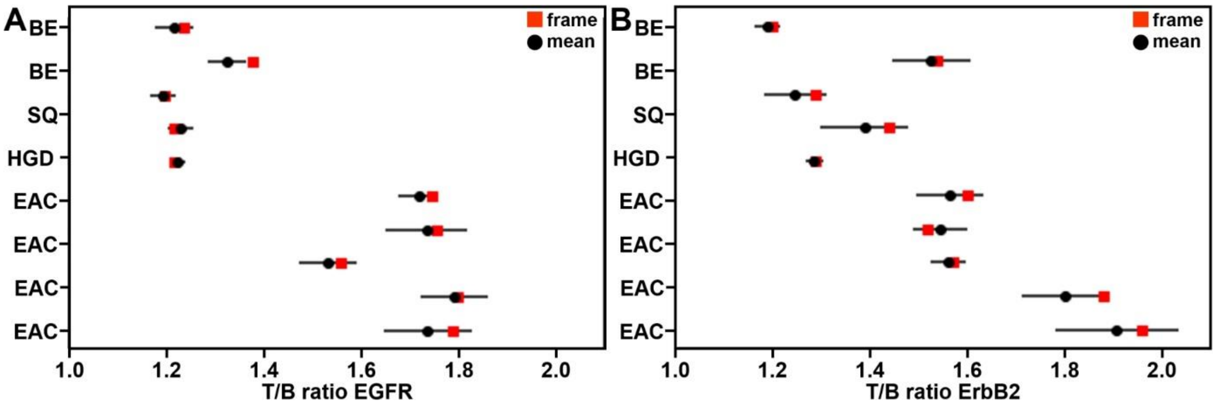


Figure 4.5 T/B variability within video segment. Multiple individual frames were extracted from representative video streams collected in vivo from the esophagus for expression of **A**) EGFR and **B**) ErbB2. The mean (oval) and standard deviation of replicate estimated T/B ratios are shown in black. The square (red) shows the value for the representative frame. For EGFR, the average coefficient of variation (CV), defined as the standard deviation divided by the mean, was 3.2%. The representative images had a 1.2% higher T/B ratio than the average of the replicates, *P* = 0.03 by paired t-test of log-transformed values. Bias was 0.78% higher in the HGD/EAC samples, 1.6% versus 0.8%, *P* = 0.45 by two-sample t-test of the differences of log-transformed data. For ErbB2, the average CV was 4.2%. Representative images had a 1.8% higher T/B ratio than the average of the replicates, *P* = 0.015. Bias was 0.02% lower in the HGD/EAC samples (1.7% versus 1.9%, *P* = 0.90).

4.4 DISCUSSION

In this chapter, we demonstrate the feasibility to detect Barrett’s neoplasia endoscopically by imaging two targets concurrently in vivo. Fluorescent labeled peptides specific for EGFR and ErbB2 were administered topically in the distal esophagus of n = 22 BE patients. These tyrosine kinases were overexpressed in progression from BE to EAC on gene expression profiles [92-94]. Two laser excitation wavelengths were delivered concurrently through a single flexible optical fiber. Adequate signal was collected by using large core, high numerical aperture fibers. The regions imaged were compared with histopathology of specimens excised via either EMR or biopsy.

To our knowledge, this is the first study to demonstrate clinical imaging of two targets at the same time during endoscopy. Many cancers, including EAC, are molecularly heterogeneous, thus detection of multiple targets is likely to be needed for accurate clinical diagnosis. Molecular biomarkers can be highly specific for disease and are expressed well before neoplastic lesions become grossly apparent. Endoscopic imaging strategies for detecting these targets in vivo can be used to guide and prioritize high risk regions for resection, reduce surveillance frequency, and minimize over diagnosis. This approach may be generalized to early cancer detection in other hollow organs.

To quantitatively evaluate the captured fluorescence images and compared to histology, frames with high contrast and minimum were selected, resulting in slightly higher T/B ratios than the mean over a 1 sec video (30 frames), Figure 4.5. However, the differences were not statistically significant. Although marginally higher T/B ratios were used, the imaging results were compared with the most advanced histological classification found on pathology. T/B ratios were calculated for both fluorescence channels. To combine the T/B ratios from the two channels, different methods were evaluated and compared to T/B from single peptide, including maximum, SVM and logistic regression. Limited by the size of the dataset, nested cross validation was used to compare models to prevent overfitting and bias. The performance of multiplexed detection, regarding to the classification accuracy and the area under the curve of the ROC curve, showed better trade off by comparison with either target alone, which indicates T/B ratio as a reliable method to quantify fluorescence imaging for detection of Barrett’s neoplasia.

However, one limitation is that the suspicious frames were manually selected for T/B calculation, which may suffer from inter and intra observer variation. Moreover, when calculating T/B ratio, only fluorescence intensity is considered. There will be morphological information for cancer detection in both fluorescence and reflectance channels. Lastly, the whole image analysis is done after the endoscopic procedure, which cannot provide real-time guidance for biopsy and surgery. A potential solution to these limitations is applying artificial intelligence (AI), which will provide real-time and automatic suspicious lesion selection and T/B calculation. In the next chapter, I will demonstrate the feasibility of real-time T/B calculation using deep learning algorithms.

Chapter 5. DEEP LEARNING BASED REAL-TIME AUTOMATIC TARGET TO BACKGROUND RATIO CALCULATION IN MULTIMODAL FLUORESCENCE ENDOSCOPY FOR ESOPHAGEAL CANCER DETECTION AND GUIDING BIOPSY

In this chapter, I propose a deep learning-based method to calculate T/B ratio automatically and in real-time, thus providing quantitative analysis for guiding biopsy and treatment.

5.1 BACKGROUND

T/B ratio has been demonstrated to provide consistent and promising results in Barrett's neoplasia detection, as shown in Chapter 2 & 4, where fluorescence molecular probes have been applied for early cancer detection in esophagus. The algorithm may be more important for more wide-spread use of fluorescence molecular probes during surveillance and screening, where a real-time and automatic computer aided diagnosis (CAD) system with T/B ratio as an indicator for cancer detection, can serve as a reliable second observer that could draw the endoscopist's eye, in real time to localize and outline suspicious lesions for biopsy. To provide this real-time lesion classification and localization system, real-time and automatic T/B calculation is needed.

This real-time T/B calculation is limited by the speed of target segmentation. Currently, Chan-Vese algorithm is applied to segment the fluorescence targets, with speed limited to only 2-3 frames/sec. The evolution to minimize the loss function in iteration constrains this algorithm to process in parallel, which makes real-time computation difficult. In addition, there will be hyperparameter adjustments for Chan-Vese algorithm when the contrast and brightness of the images change. The performance of the segmentation is greatly depended on the quality of the image, which makes automatic T/B calculation more difficult. Moreover, to achieve this automatic CAD system, an automatic frame selection is needed to reject frames with diverse image artifacts, including (1) saturation, (2) artefact, (3) blur, (4) dye pooling, (5) bubble and (6) instrument (mother endoscope) and to select frames with suspicious regions for further T/B quantification, which is currently done manually in our previous study (Chapter 4).

Deep learning, a machine learning subfield, has been successfully applied to many areas of science and technology, including computer vision [101, 102], speech recognition[103, 104], games [105], bioinformatics [106] etc. and has been expanded to the field of medical image analysis. There have been immense researches in computer-aided detection and diagnosis using deep learning techniques in various imaging modalities, e.g. MRI, computed tomography, ultrasound etc. [107, 108]. The main contribution of deep learning in medical image analysis is mainly around two topics: (1) detection and classification, (2) localization and segmentation.

- Detection and classification. Convolutional neural network (CNN) has become state-of-the-art tool for medical image classification, which has been widely used for cancer classification. Litjens et al. used CNN to detect cancer areas from H&E-stained whole slides for prostate and breast cancer detection. All slides containing prostate cancer and metastases of breast cancer could be identified automatically while 30–40% of the slides containing benign and normal tissue could be excluded without any human intervention [109]. Ardila et al. proposed an end-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, and achieved a state-of-the-art performance (94.4% area under the curve) on 6,716 National Lung Cancer Screening Trial cases [110]. Kooi et al. showed that a CNN model trained on a large data set of mammographic lesions (45000 images) outperformed a state-of-the art CAD system trained using manually designed features (AUC: 92.9% vs. 90.6%) [111].

- Localization and segmentation. Another contribution of deep learning to medical image field is the potential to improve the speed and accuracy of cancer localization and segmentation in clinical settings. A CNN based algorithm was applied for automatic localization and segmentation of rectal cancers on multiparametric MR imaging, with segmentation results comparable to those manually labeled by experts (DSC 0.70) [112]. A hybrid network that fuses DenseNet [113] and UNet [114] together, achieved competitive performance for liver and tumor segmentation from 3D CT volumes. Both features from inter- and intra-slices were extracted and combined for better 3D segmentation [115]. A two-pathway CNN, that extracted both local and global context, was applied to segment brain tumor for MRI scan and improved the accuracy and speed over other algorithms reported on the 2013 BRATS [116].

There are also increasing number of studies that investigated deep learning applications to endoscopic imaging [117 - 120]. With the development of massive parallel architecture, Graphic Processing Unites (GPUs), deep learning techniques have demonstrated promising results in real time endoscopic cancer surveillance [117, 118]. Wang et al. developed a deep-learning algorithm by training on 5,545 colonoscopy images from 1,290 patients for polyp detection and localization, with processing speed 25 frames per second. The algorithm achieved a 94.38% sensitivity and 95.92% specificity validated on 27,113 colonoscopy images [117]. The robustness and real-time capability of deep learning allows our suspicious frame selection system with real-time and automatic T/B calculation. Although the segmentation of the fluorescence images remains challenging due to the “patchy” appearance of the fluorescence targets, non-uniform intensity in both targets and background, and loss in contrast and resolution by the scattering of light from biological tissue, there have been deep learning models showing promising results for nucleus segmentation in fluorescence microscopy [121, 122]. To the best of our knowledge, there is no deep learning based method for wide-field fluorescence endoscopy (1) to provide real-time suspicious frame selection, and (2) to provide real-time and automatic quantification using T/B ratio.

In this chapter, we describe a deep learning based method (CNN) to automatically select suspicious video frames and reject frames with artefacts. In addition, we develop an algorithm for automatic and real -time T/B calculation. Different deep learning architectures are evaluated to segment fluorescence targets, including BiSeNet [123] and UNet [114]. We validated the whole system in two clinical studies. The CNN model for frame selection achieved uniform and consistent high performance with a sensitivity of 96.4%, a specificity of 96.6%, a precision of 95.5% and a AUC 0.989, and with a processing speed 20ms on a 480 * 480 image. For the segmentation task, the best model achieved mIOU 0.89 and 0.88, respectively for the multiplex and dimer study with processing speed 36ms on a 480* 480 image.

5.2 METHOD

5.2.1 *Convolutional Neural Networks*

We trained different CNN architectures as backbone to extract features for the following tasks, (1) classification and (2) segmentation. These CNN architectures usually consist of the following layers:

- **Convolutional layers** consist of a set of learnable filters, which slide across the width and height of the input volume. Dot products are then computed between the entries of the filter and the portion of the image over which the filter is hovering. Unlike traditional image feature extraction methods where filters are manually defined (e.g. Gabor filter bank), convolutional layers learn the weights and biases of individual feature maps, which provide customized, task-specific feature extractors. The entries of the filter can be optimized to minimize the loss function using methods, such as stochastic gradient descent, where gradients are computed efficiently via the backpropagation algorithm. The network will learn to activate visual features such as edges of different orientations on the first few layers, and eventually the entire shape on higher layers.
- **Batch normalization layers** are introduced to standardize the inputs of each layer for each mini batch, which have been shown to stabilize the learning process and dramatically reduce the number of training epochs required to train deep networks. More specifically, batch normalization normalizes the distribution of each layer's inputs to zero mean and unit standard deviation, to reduce the effects of covariance shift [125].
- **Activation functions** are applied to the output of every layer to introduce non-linearity to the network, thus helping the network learn complex patterns of the image. Without these non-linear activation functions, a neural network would be mathematically equivalent to a linear regression model. Common activation functions are Sigmoid, Tanh, ReLU, Leaky ReLU. Notable activation functions are Swish, Hard-Swish, which tend to perform better than traditional activation functions on deeper models [126].
- **Pooling layers** reduce the spatial size of the convoluted features, thus decreasing the number of parameters and computation power required in the network. The outputs of

pooling layers summarize information from neighborhood regions and extract dominant features for training. There are mainly two types of pooling: max pooling and average pooling. Max pooling returns the maximum from the region covered by the filters. On the other hand, average pooling returns the average from the region covered by the filters. The most common form is a pooling layer with a filter size 2x2, which discards 75% of the activations.

- **Skip connections** add the outputs from the previous layer to the outputs of adjacent layer, which allows the gradient to flow from the first layer to the last layer during backpropagation, thus preventing the vanishing gradient problem [127]. Skip connections enable feature reusability and stabilize training process.

All these layers are assembled to form different backbone architectures for feature extraction. In this study, two different backbone architectures, MobileNetV2 [128] and Xception [129] were used.

MobileNetV2 is a light-weight model, where depthwise separable convolution is introduced to dramatically reduce the size of the model and the computation cost. With an inverted residual structure, MobileNetV2 achieves state-of-the-art performances for classification, object detection and semantic segmentation. The convolutional blocks of MobileNetV2 and the overall architecture are shown in Figure 5.1 and Table 5.1, respectively [128].

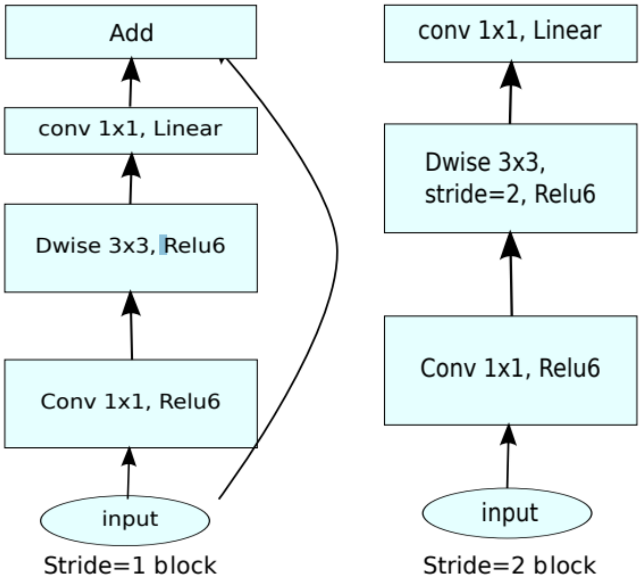
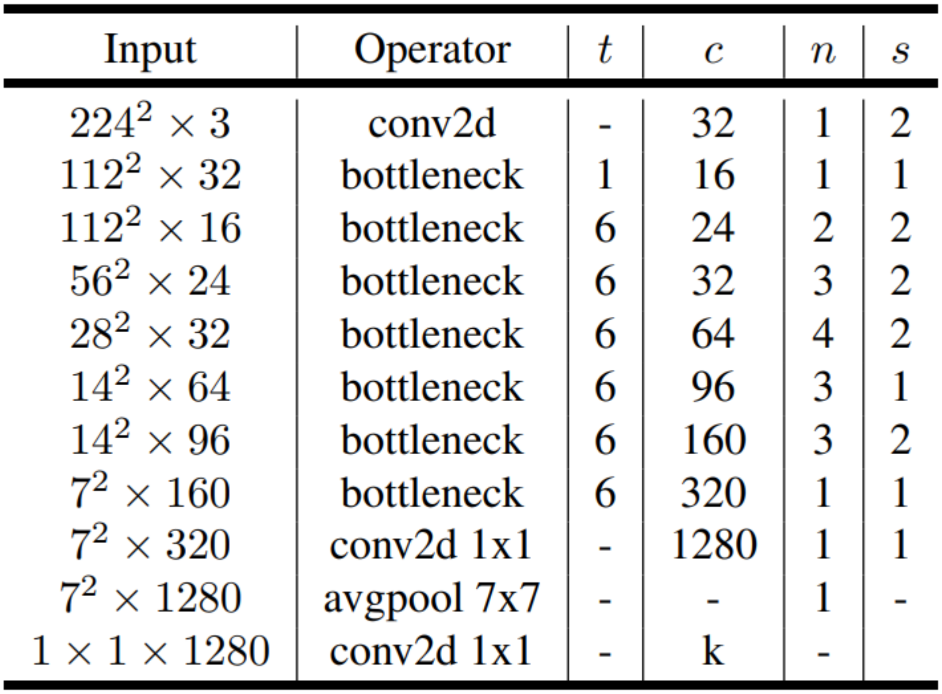


Figure 5.1 MobileNetV2 convolutional blocks [128]

Table 5.1 MobileNetV2 overall architecture [128]



Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Notes: Each line describes a sequence of 1 or more identical (modulo stride) layers, repeated n times. All layers in the same sequence have the same number c of output channels. The first layer of each sequence has a stride s and all others use stride 1. All spatial convolutions use 3 × 3 kernels. The expansion factor t is always applied to the input size

Xception is a deep convolutional neural network architecture with Depthwise Separable Convolution applied. The overall architecture is shown in Figure 5.2.

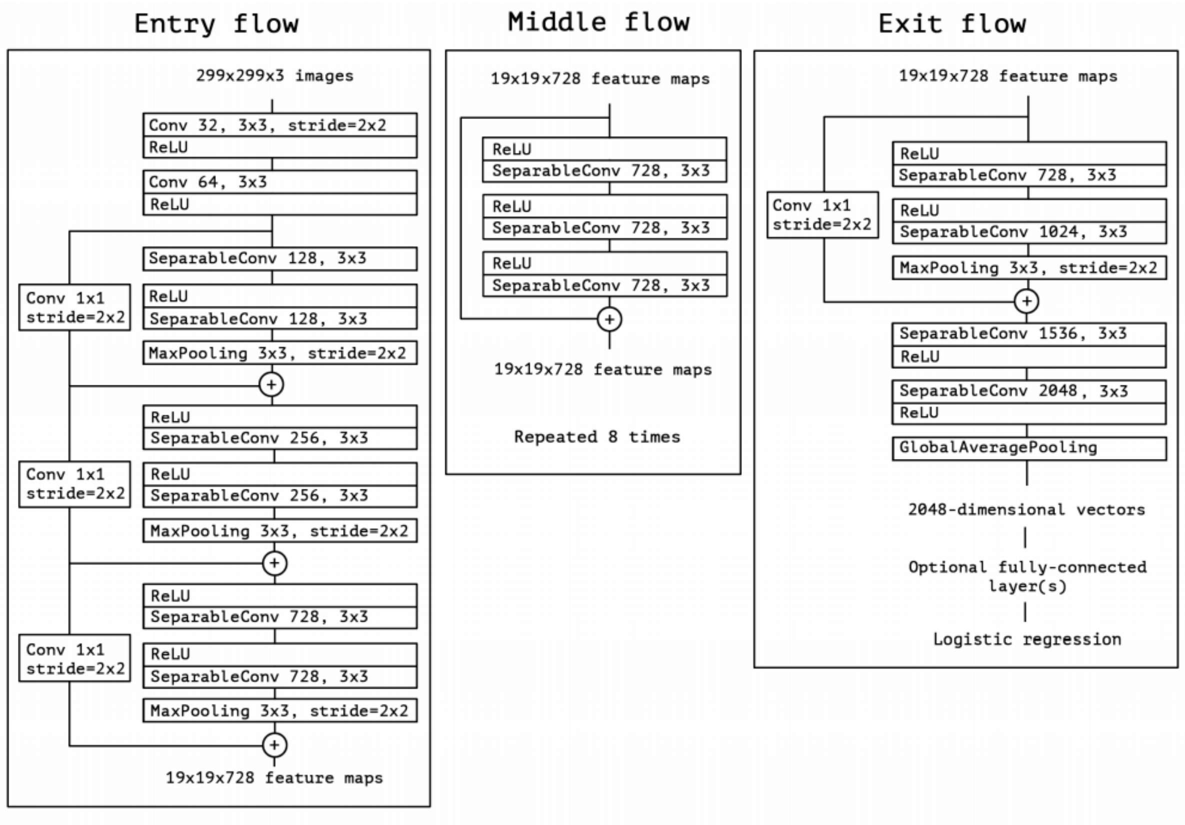


Figure 5.2 Xception overall architecrure [129]

5.2.2 *Suspicious frame selection model*

A CNN architecture with Xception or MobileNetV2 backbone, followed by a global average pooling layer and 3 fully connected layers, was applied for suspicious frame selection.

5.2.3 *Semantic segmentation model*

Architectures for semantic segmentation usually consist of two parts: encoder and decoder. The encoder compresses the input image into a smaller vector and extracts context information from it. On the other head, the decoder expanded the extracted information and reconstructed an output with the size of the original input image to get a pixel-level classification. In this study, 2 different architectures were evaluated and compared for fluorescence target segmentation, including BiSeNet and UNet.

BiSeNet [123] contains two pathways, a Spatial Path with a small stride to preserve the spatial information and generate high-resolution features, and a Context Path as encoder to obtain sufficient receptive field and to extract context information for segmentation. A Feature Fusion Module is introduced to combine features from two pathway efficiently. Xception and MobileNetV2 were used as backbone to extract features within Context Path. A global average pooling was applied on the tail of these lightweight models to provide the maximum receptive field with global context information. An Attention Refinement Module (ARM) was introduced to refine the features of last two stages of the Context Path. The overall architecture of BiSeNet is shown in Figure 5.3.

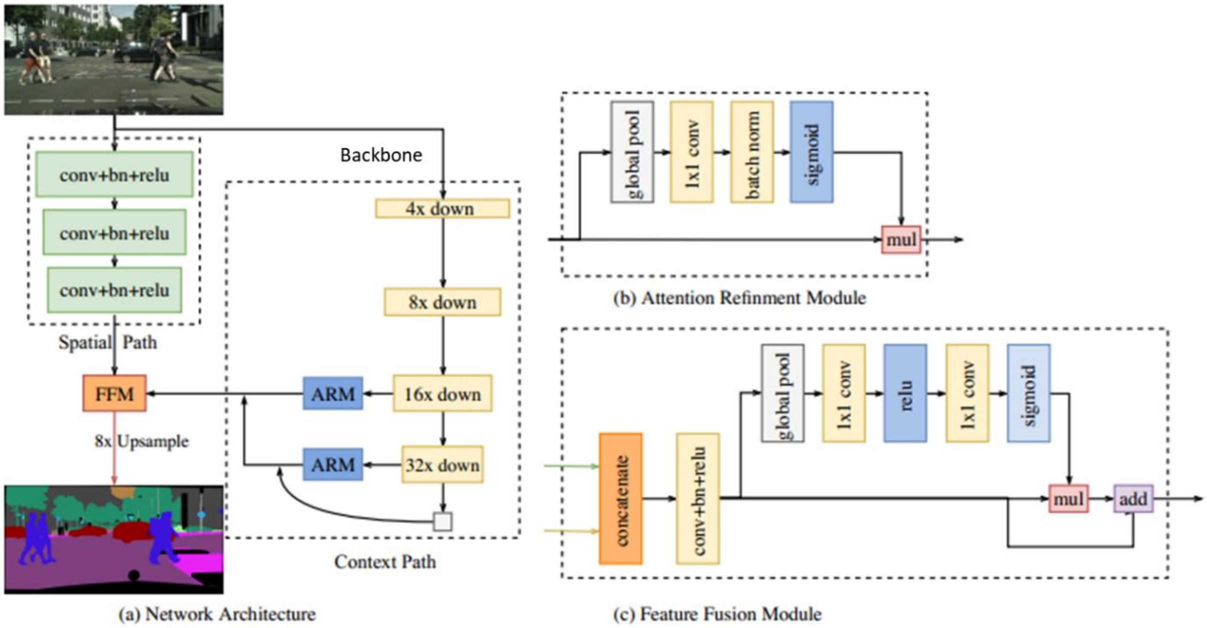


Figure 5.3 An overview of the Bilateral Segmentation Network. (a) Network Architecture. (b) Components of the Attention Refinement Module (ARM). (c) Components of the Feature Fusion Module (FFM) [123].

UNet consists of a symmetric encoder and decoder. The encoder extracts features from images. The decoder concatenates high-resolution spatial information with up-convoluted context information. Figure 5.4 shows a conceptual UNet architecture. In this study, Xception and MobileNetV2 were used as backbone to extract features for encoder.

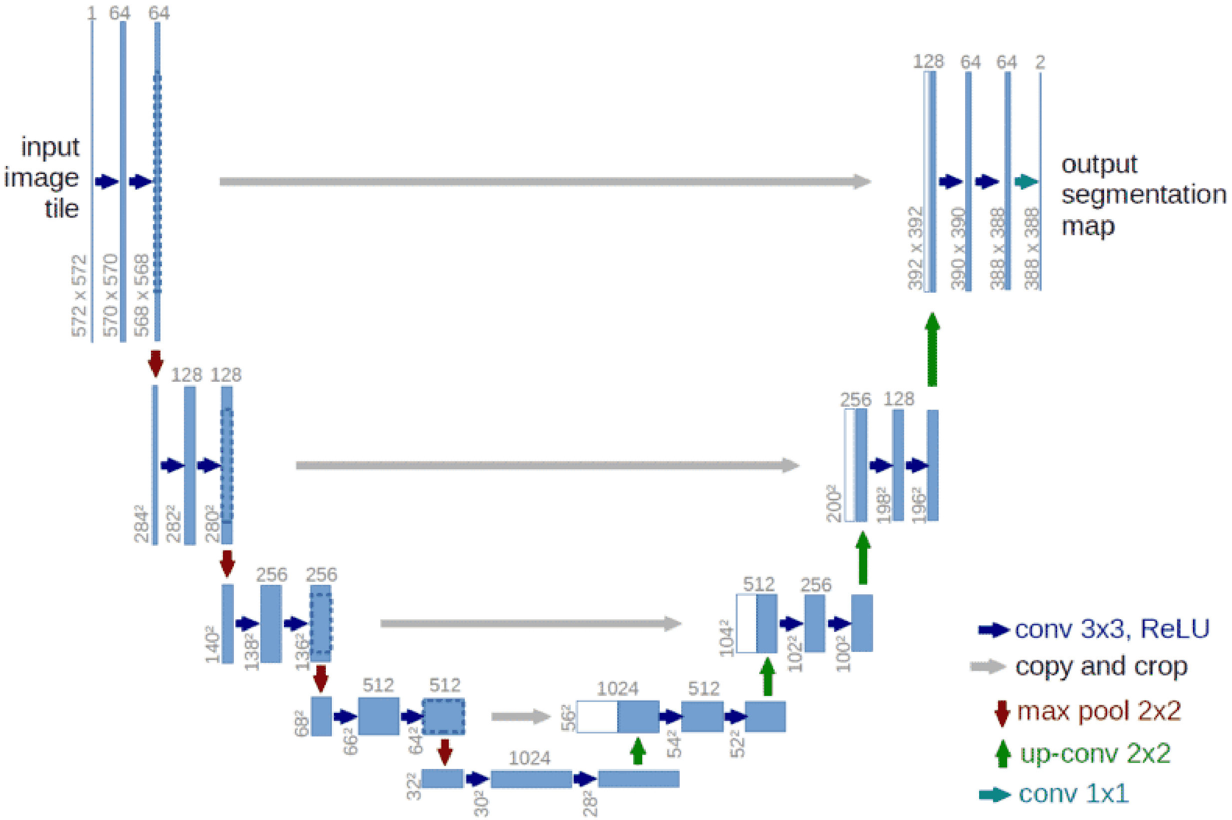


Figure 5.4 UNet architecture [114]

All experiments were implemented using the Tensorflow [130] software libraries.

5.2.4 *Model regularization*

Different training techniques were applied to decrease overfitting and to improve performance without sacrificing the training cost. In this study, data augmentation was applied to increase the variability of the input images, so that the trained model could handle images from different environments. We used 3 different types of image transformations for augmentation: (1) geometric distortions (2) brightness and contrast adjustment (3) blur and noise. For geometric distortions, we used random scaling, shifting, flipping and, rotating in the full range of 0 – 90 degrees. For blur and noise, we used random motion blur, Gaussian blur, sharpen, and Gaussian noise. All the augmentation methods mentioned above are pixel-wise adjustment, which retained the original pixel information. Another technique we used was "early stopping", where a subset of the training data is reserved as validation data for monitoring accuracy during the training process. The weight with maximum validation accuracy was obtained. In addition, we used the weights pre-trained on the ImageNet data corpus as initialization for our backbone architectures. The ImageNet challenge dataset contains 1.2 million natural images, such as cars, dogs etc., which shares fundamental features that can be transferred to medical images.

5.3 EXPERIMENTS

5.3.1 *Datasets and preprocessing*

Two datasets from two clinical studies were used for training and/or testing deep learning models presented in this work:

1. The set contains 15 videos from patients with history of Barrett's esophagus or Barrett's neoplasia (HGD/EAC) and recruited for multiplex imaging clinical trial. Fluorescence labeled peptides QRH*-Cy5 and KSP*-IRDye800, targeting EGFR and Her2, were topically applied and videos were recorded using mmSFE with a resolution 720 * 720. Separate channels were used to record fluorescence from QRH*-Cy5 (red) and KSP*-IRDye800 (green), and reflectance (blue). The combined duration of 15 videos was approximately 96 mins (~2900 frames).

- A separate set contains 20 videos from patients with history of Barrett's esophagus or Barrett's neoplasia (HGD/EAC) and recruited for dimer study. Single IRDye800 labeled peptide targeting both EGFR and Her2 was topically applied. The NIR fluorescence was projected to green channel, with reflectance as blue channel. The combined duration of 20 videos was approximately 88 mins (~2650 frames)

All images from the two datasets were preprocessed to have the same formats, before being passed to the deep learning models. First, the individual frames were rescaled to a fixed size of 480 * 480. Then, the value of the pixels in each frame was normalized by dividing 255.0. For suspicious frame selection task, each image from multiplex study was separated into two images, with green for fluorescence and blue for reflectance, to match the images from dimer study. For fluorescence target segmentation task, the fluorescence channels from multiplex images were extracted separately as grayscale images. Similarly, single fluorescence channel from dimer images was extracted for the segmentation task.

5.3.2 *Model development*

We trained and evaluated suspicious frame selection model and fluorescence target segmentation model separately.

5.3.2.1 Frame selection model

we used ImageNet pretrained MobileNetV2 and Xception as backbone. A softmax layer was applied to classify extracted features into two classes: 1) frames containing bright suspicious lesions with minimum motion blur, 2) frames containing normal esophagus with no bright suspicious lesions, and frames with different artifacts, including dye pooling, air bubble, saturation, blur, instrument (mother endoscope). The development dataset from the two clinical studies were divided into two parts.

- Training and validation. 3427 images were extracted from 15 videos from multiplex study and 10 videos from dimer study. There are 1475 images with suspicious bright regions and

1952 images with different artifacts. The 3427 images were then stratified into training (83%) and validation (17%).

- Testing. 1039 images were extracted from the remaining 10 videos from dimer study, which contains 443 images with suspicious bright regions and 596 images with artifacts.

We augmented the dataset by using different combinations of image transformations during training. We adjusted the batch size and optimized the initial learning rate for each backbone architecture individually. We employed the same learning rate schedule (decreasing by 0.8 every 5 epochs). The neural networks were trained with the ADAM optimizer [131] to minimize the binary crossentropy loss. In terms of evaluation, accuracy and AUC was computed and compared for different combinations of backbone and augmentation techniques over test set.

5.3.2.2 Fluorescence target segmentation model

Different model architectures BiSeNet and UNet, with backbone MobileNetV2 or Xception, were trained and evaluated for fluorescence segmentation. Only frames with suspicious regions were selected from the two clinical studies and divided into three sets.

- Training and validation. 1006 images were extracted from 11 videos from multiplex study. The images were separated into training (960) and validation (45).
- Testing A. 100 images were extracted from 7 videos from multiplex study with different suspicious regions from training and validation set.
- Testing B. 100 images were extracted from dimer study.

All the images were paired with pixel-level annotated segmentation map, which contains one foreground and one background. The annotated segmentation maps were generated using Chan-Vese algorithm described in Chapter 2. We applied the same training protocol for training the frame selection model. The batch size and optimized the initial learning rate were adjusted for each model and backbone individually. ADAM optimizer was used to minimize the sparse categorical crossentropy. The performance is measured in terms of overall pixel-wise accuracy and pixel intersection-over-union averaged across the 2 classes (mIOU). IOU is also known as Jaccard index J (Equation 5.1).

$$IOU\ (A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Equation 5.1

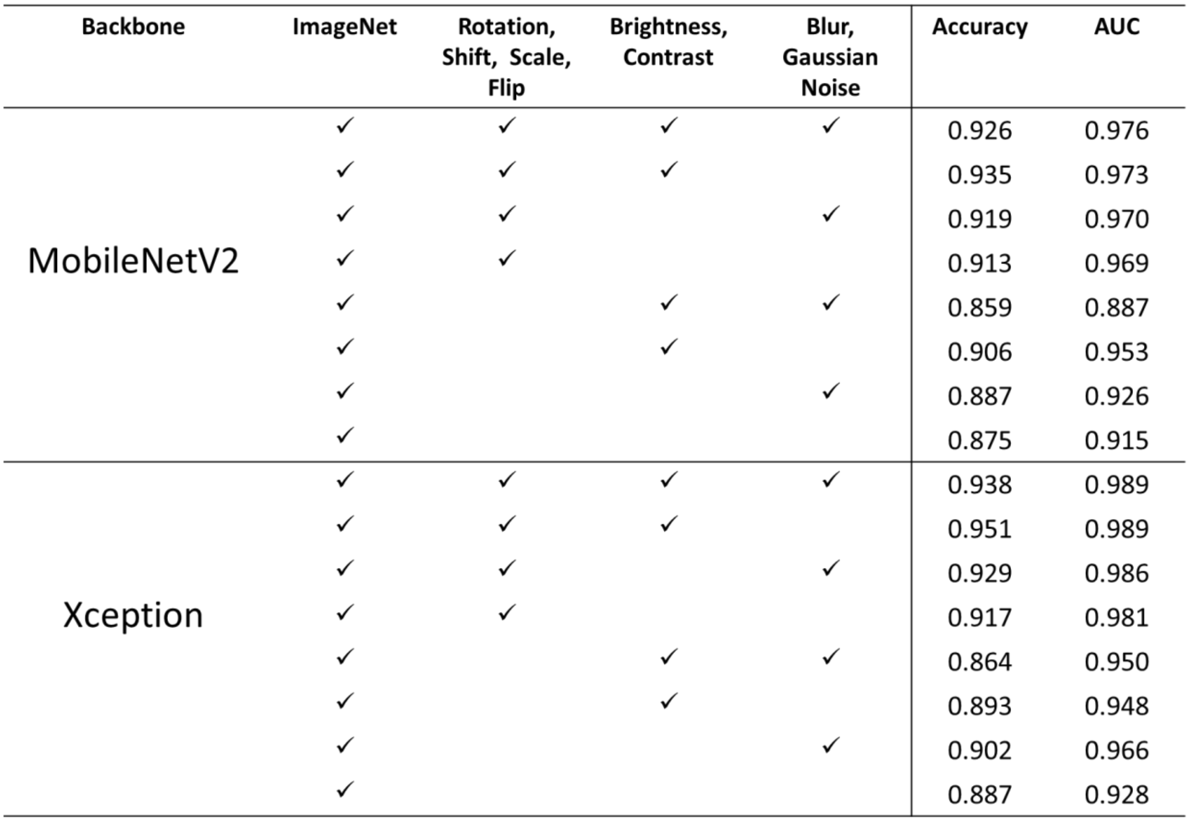
Note: (A) ground truth (B) segmented fluorescence target

5.4 RESULTS

5.4.1 *Suspicious frame selection*

Suspicious frame selection results are shown in Table 5.2. Adding augmentation during training improves network performance. Networks with geometric distortions surpassed those without geometric distortions by a significant margin. We also observed that adjusting the brightness and contrast of the image during training may improve the performance of the model. Xception backbone performs slightly better than MobileNetV2

Table 5.2 Summary of suspicious frame selection results for different architecture (MobileNetV2 & Xception) and different combinations of augmentation



Backbone	ImageNet	Rotation, Shift, Scale, Flip	Brightness, Contrast	Blur, Gaussian Noise	Accuracy	AUC
MobileNetV2	✓	✓	✓	✓	0.926	0.976
	✓	✓	✓		0.935	0.973
	✓	✓		✓	0.919	0.970
	✓	✓			0.913	0.969
	✓		✓	✓	0.859	0.887
	✓		✓		0.906	0.953
	✓			✓	0.887	0.926
	✓				0.875	0.915
Xception	✓	✓	✓	✓	0.938	0.989
	✓	✓	✓		0.951	0.989
	✓	✓		✓	0.929	0.986
	✓	✓			0.917	0.981
	✓		✓	✓	0.864	0.950
	✓		✓		0.893	0.948
	✓			✓	0.902	0.966
	✓				0.887	0.928

At a sensitivity level (true-positive rate) of 96.4%, the best model (Xception backbone with geometric distortions, brightness, and contrast adjustment augmentation during training) had a specificity of 96.6% and a precision of 95.5%. The confusion matrix and the ROC curve of the test data are shown in Figure 5.5. As can be seen, there is a reasonably large range of high sensitivities, high specificity, and high precision (all greater than 90%).

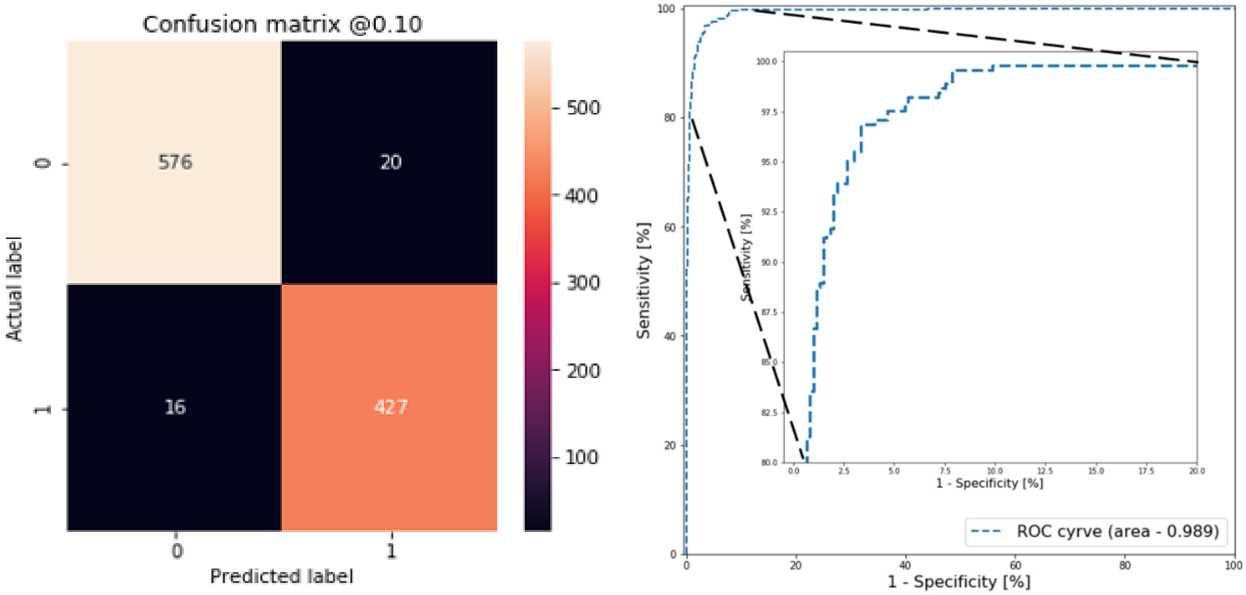


Figure 5.5 The confusion matrix and ROC curve for the best model (Xception backbone with geometric distortions, brightness, and contrast adjustment augmentation during training), evaluated over test data

5.4.2 *Fluorescence target segmentation*

1006 images were used to train and validate the model on the segmentation task. Results over the test sets, as presented in Table 5.3 and Table 5.4, showed that

(1) Networks with geometric distortions augmentation during training surpassed those without geometric augmentation by a significant margin, which is consistent with our findings for frame selection model.

(2) Neither of (1) adjusting the brightness and contrast of the image and (2) adding noise or blur to the image during training showed an improvement in performance.

(3) UNet outperformed BiSeNet in segmentation accuracy.

(4) Using pretrained ImageNet weights as initialization did not improve the performance.

(5) Xception backbone performs slightly better than MobileNetV2 for BiSeNet, but not for UNet.

Table 5.3 Summary of fluorescence segmentation results for BiSeNet model

Model + Backbone	ImageNet	Rotation, Shift, Scale, Flip	Brightness, Contrast	Blur, Gaussian Noise	Multiplex		Dimer	
					Accuracy	mIOU	Accuracy	mIOU
BiSeNet + MobileNetV2	✓	✓	✓	✓	0.977	0.868	0.959	0.844
	✓	✓	✓		0.977	0.869	0.959	0.849
	✓	✓		✓	0.976	0.864	0.962	0.858
	✓	✓			0.973	0.848	0.958	0.846
	✓		✓	✓	0.972	0.839	0.954	0.824
	✓		✓		0.970	0.828	0.952	0.824
	✓			✓	0.976	0.864	0.956	0.832
	✓				0.967	0.818	0.952	0.818
		✓	✓	✓	0.977	0.871	0.957	0.841
		✓	✓		0.977	0.870	0.959	0.846
		✓		✓	0.978	0.874	0.964	0.865
		✓			0.977	0.872	0.962	0.858
			✓	✓	0.968	0.819	0.948	0.805
			✓		0.968	0.826	0.955	0.831
				✓	0.971	0.836	0.952	0.817
					0.969	0.828	0.954	0.829
BiSeNet + Xception	✓	✓	✓	✓	0.978	0.873	0.964	0.864
	✓	✓	✓		0.978	0.874	0.962	0.859
	✓	✓		✓	0.978	0.878	0.966	0.873
	✓	✓			0.980	0.886	0.965	0.868
	✓		✓	✓	0.973	0.841	0.953	0.821
	✓		✓		0.974	0.850	0.954	0.829
	✓			✓	0.975	0.856	0.957	0.837
	✓				0.975	0.859	0.955	0.831
		✓	✓	✓	0.978	0.873	0.963	0.861
		✓	✓		0.978	0.875	0.962	0.859
		✓		✓	0.980	0.885	0.964	0.865
		✓			0.979	0.880	0.965	0.870
			✓	✓	0.970	0.827	0.956	0.830
			✓		0.971	0.833	0.951	0.815
				✓	0.972	0.840	0.956	0.834
					0.972	0.845	0.956	0.833

Table 5.4 Summary of fluorescence segmentation results for UNet model

Model + Backbone	ImageNet	Rotation, Shift, Scale, Flip	Brightness, Contrast	Blur, Gaussian Noise	Multiplex		Dimer	
					Accuracy	mIOU	Accuracy	mIOU
UNet + MobileNetV2	✓	✓	✓	✓	0.980	0.889	0.962	0.857
	✓	✓	✓		0.982	0.897	0.962	0.862
	✓	✓		✓	0.981	0.891	0.968	0.880
	✓	✓			0.979	0.881	0.963	0.863
	✓		✓	✓	0.974	0.847	0.955	0.829
	✓		✓		0.973	0.845	0.955	0.830
	✓			✓	0.976	0.865	0.959	0.845
	✓				0.970	0.829	0.952	0.816
		✓	✓	✓	0.979	0.880	0.963	0.865
		✓	✓		0.979	0.877	0.963	0.860
		✓		✓	0.979	0.882	0.967	0.877
		✓			0.979	0.881	0.964	0.865
			✓	✓	0.972	0.839	0.949	0.810
			✓		0.970	0.834	0.962	0.857
				✓	0.972	0.840	0.958	0.844
					0.960	0.797	0.957	0.839
UNet + Xception	✓	✓	✓	✓	0.982	0.897	0.966	0.872
	✓	✓	✓		0.979	0.883	0.965	0.870
	✓	✓		✓	0.981	0.892	0.966	0.875
	✓	✓			0.978	0.877	0.964	0.867
	✓		✓	✓	0.967	0.823	0.954	0.830
	✓		✓		0.972	0.840	0.957	0.838
	✓			✓	0.976	0.862	0.957	0.838
	✓				0.972	0.845	0.956	0.835
		✓	✓	✓	0.981	0.892	0.964	0.866
		✓	✓		0.981	0.893	0.966	0.875
		✓		✓	0.981	0.895	0.966	0.875
		✓			0.981	0.891	0.966	0.875
			✓	✓	0.969	0.831	0.960	0.852
			✓		0.971	0.841	0.958	0.844
				✓	0.973	0.851	0.960	0.850
					0.972	0.842	0.957	0.837

Note: mIOU, mean IOU over two classes

UNet with Xception or MobileNetV2 backbone and geometric distortions augmentation performed better at segmenting the fluorescence images. UNet with (1) MobileNetV2 backbone, (2) ImageNet pretrained weight initialization, (3) geometric distortion, blur and noise augmentation achieved the highest mIOU and overall pixel-wise accuracy for both test datasets: mIOU = 0.891, accuracy = 0.981 for multiplex dataset and mIOU = 0.880, accuracy = 0.968 for dimer dataset. Figure 5.6 shows examples of segmentation results tested on the two datasets. Deep learning models performed similarly in segmenting the fluorescence targets compared to Chan-Vese algorithm but processed at least 10 times faster.

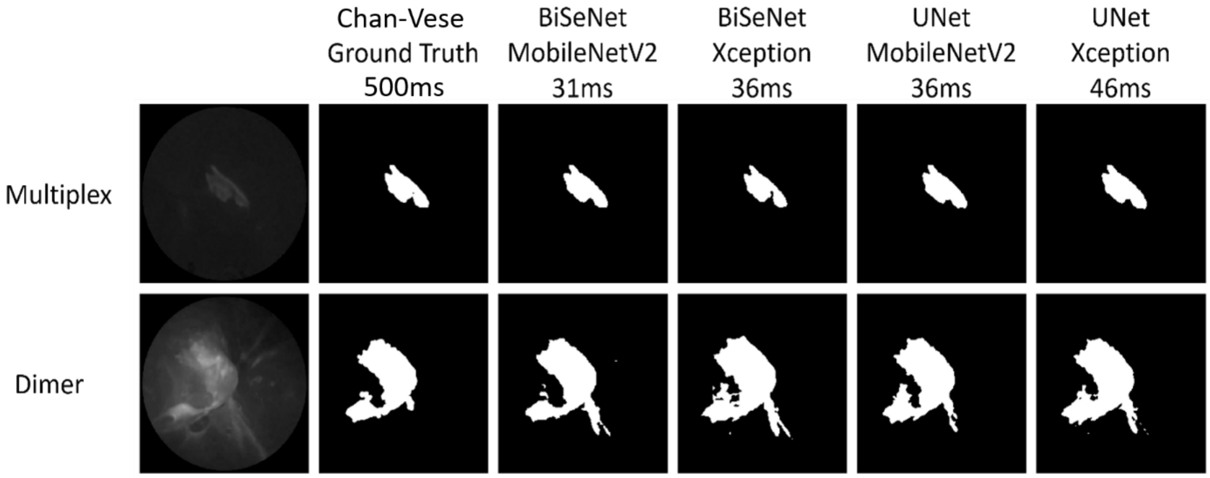


Figure 5.6 Representatives of segmentation results for BiSeNet and UNet model with MobileNetV2 and Xception backbones

5.4.3 *Speed*

We achieved an average 20ms processing time for the frame selection task and 31~46ms processing time for the segmentation task depending on the model architecture achieved, on a NVIDIA RTX 2080S external GPU. The total processing time for each image frame, which is equivalent to the total time of the two tasks, was 50 ~ 75ms.

5.5 DISCUSSION

In this chapter, I presented a deep learning-based pipeline to calculate T/B ratio automatically and in real-time, which consisted of two tasks, (1) frame selection and (2) fluorescence segmentation. A CNN model was trained for frame selection with processing speed 20ms on a 480 * 480 image. Frames with suspicious bright regions were selected, while frames with artifacts, including dye pooling, bubble, saturation, blur etc., were rejected. The model achieved uniform and consistent high performance with a sensitivity of 96.4%, a specificity of 96.6%, a precision of 95.5% and a AUC 0.989. For the segmentation task, UNet and BiSeNet with Xception and MobileNetV2 backbones were evaluated on the two datasets from two different clinical studies. The best model achieved mIOU 0.89 and 0.88, respectively for the multiplex and dimer study with processing speed 36ms on a 480* 480 image. The system can achieve 50 frame per second for frame selection task only. Once the frame is selected for T/B calculation, additional time is needed for the segmentation, which will drop the total speed to ~15 frame per second. A real-time CAD software should operate at a speed of at least 25 frames per second according to PAL and NTSC standards, where video encodings are standardized to 25 or 30 frames per second. To accelerate the speed to 30 frames per second. we can implement the algorithm in a multi-threaded structure, 1 thread for frame selection and 1 thread for segmentation. To further accelerate the speed, we can downsample the input to a smaller size, e.g. 320*320, which will reduce the number of operations by half and double the speed.

One limitation of this study is that only 35 videos from 35 patient were applied for training and testing the models, which limited the robustness of the obtained model. However, instead of classifying each frame as HGD/EAC or normal, a simpler task was investigated to select frames with suspicious bright regions for further T/B calculation. For the segmentation task, the single foreground object class (fluorescence) and the high tolerance of segmentation accuracy diminish the effect of a small dataset. In addition, different training techniques were applied to further minimize the effect of a small dataset, including image augmentation, ImageNet pretrained weight initialization. To further improve the robustness of the model, synthetic images can be generated for training. These synthetic images can be generated using Monto-Carlo modelling with optical ray-tracing technique. Carles reports a hybrid Monte-Carlo optical ray-tracing technique for modelling of complete imaging systems of arbitrary biological structures embedded within complex scattering media [124]. It is also feasible to synthesize these photo-realistic images using deep learning techniques, especially generative adversarial networks (GANs) [132]. Shin et al. trained a GAN based on pix2pix to generate synthetic abnormal MRI images with brain tumors [133]. These synthesized images can serve as a form of data augmentation and provide large amount of data for deep learning training in medical imaging field, where collecting patient data is costly and challenging.

Overall, this automatic CAD system using deep learning will be ideal for real-time guiding biopsy and surgery with T/B ratio as an indicator for cancer detection. Especially, the combination of fluorescence molecular probes and artificial intelligence-based image quantification can greatly improve routine screening or surveillance, where an automatic CAD system can serve as an reliable second observer to localize and outline suspicious lesions highlighted by fluorescence probes, thus providing real-time guidance for biopsy.

Chapter 6. DISCUSSION AND CONCLUSION

This chapter summarizes the contribution of multimodal fluorescence imaging quantification methodology in this dissertation and the impact in cancer detection, guiding biopsy and image guided surgery. The overall outcomes of the studies are described, the distinguished contributions compared to previous literature are summarized, the limitations are discussed, and future directions are motivated.

6.1 OUTCOMES AND CONTRIBUTIONS

This work developed quantitative analysis algorithms for multimodal fluorescence imaging, thus assisting with cancer detection, image guided biopsy and image guided surgery.

Methodology-wise, an automatic T/B ratio calculation pipeline has been first developed (Chapter 2) to evaluate fluorescence molecular endoscopy for early cancer detection in esophagus. This T/B ratio provides a relative quantification for the amount of molecular probe binding in the target region compared to the surrounding tissue (background), and indicates the relative level of differentially expressed biomarkers associated with cancer progression, under assumptions:

- The molecular probes are uniformly bound to the biomarkers.
- The fluorophore is uniformly excited. With a fixed quantum yield and a calibrated imaging system, the detected fluorescence intensity is linear to the number of the expressed biomarkers.
- There is no fluorescence intensity saturation.
- The background fluorescence is from molecular probes bound to normal cells. The detected background fluorescence is greater than system noise level.

- There is little autofluorescence detected in NIR range.

Since cancer progression is associated with biomarker overexpression, T/B ratio can be diagnostic for cancer detection. To achieve this quantitative T/B calculation, the fluorescence targets are segmented using the Chan-Vese algorithm, which is more robust in noisy low-contrast fluorescence images. We extend the use of T/B quantification to allow for real-time display of the highlighted regions. This quasi-real-time T/B calculation can be embedded into fluorescence endoscope systems, e.g. mmSFE, to provide quantitative evaluation for guiding biopsy and early cancer detection. This work has been published in Journal of Medical Imaging.

To further improve the T/B calculation framework, deep learning algorithms are applied for automatic frame selection and fluorescence target segmentation (Chapter 5). With the support of GPU, the frame selection and segmentation model can run at least 30 frame per second. This real-time and automatic T/B calculation will be suitable for routine screening or surveillance using capsule endoscopy, and especially in under-served/under-developed areas where there is a lack of experienced clinicians.

In Chapter 3, a novel framework is developed to calibrate NIR fluorescence to the ground truth, the histological images, thus evaluating fluorescent agents in delineating tumor cores and margins. This pipeline applies multiple fluorescence quantitation methods: distance compensation, specular/ambient light rejection. The calibrated fluorescence images are then registered to histological slices, using an optical attenuation model and average intensity projection. Unlike methods from other studies, where selected best histological slices are compared to fluorescence images, this multi-step registration method uses stacked H&E slides from surface to deeper tissue, to generate a synthesized 2D tumor map, which accounts for the variations in tissue optical properties and tumor depth. This work has been published in Journal of Medical Imaging.

In terms of applications of our method, T/B ratio have been applied in multiple clinical trials with single or multiple fluorescence labeled peptides (Chapter 2 & 4). In Chapter 4, T/B ratio is applied to quantify fluorescence from 2 molecular probes specific for EGFR and ErbB2, for esophageal neoplasia detection. This work has been submitted to Gut. Although this T/B ratio has been initially investigated for the early cancer detection and guiding biopsies for patients with Barrett's esophagus, yet same technique can also be transferred to the diagnosis and/or intervention in other parts of the body, such as colon, bile duct and image-guided brain surgery, using fluorescence molecular endoscopy (mmSFE).

In summary, this dissertation addresses challenges in multimodal fluorescence imaging analysis, such as calibration of fluorescence images, correlation to histology for both classification and pixel-wise registration, and use of ratiometric methods for quantification. With an accurate quantification, multimodal fluorescence imaging could potentially increase the accuracy of early cancer detection, thus improving patient survival outcomes. This dissertation work also led the development of an advanced fluorescence endoscopy into a quantitative and computer-aided supporting system.

6.2 LIMITATION AND FUTURE WORK

The methodologies developed can be used in many ways beyond those presented in this dissertation. The body of work can be also extended with the development of advanced imaging system and computer vision techniques. The limitations of this work and future applications are discussed below.

6.2.1 *Dye administration and tumor type impacts on T/B ratio*

T/B ratio shows promising results in quantifying multimodal fluorescence imaging for esophageal cancer detection. However, we only evaluate this T/B method on topically applied contrast agents and shallow tumor type (esophageal cancer). The performance of T/B ratio on Intravenous (IV) injected contrast agents and other tumor types, e.g. infiltrative brain tumor, need to be re-evaluated. Figure 6.1 shows representative images from de Jongh et al. study, where C-Met targeted fluorescence contrast agents were either topically applied or IV injected for Barrett's neoplasia detection [134]. As can be seen, a more homogenous dye distribution with stronger background fluorescence can be achieved after IV administration; while topically application enables higher local dye contrast, which results in a higher T/B ratio. When calculating T/B ratio using IV injected contrast agents for cancer classification, we expect a smaller T/B threshold. In addition, T/B ratio would perform better for shallow tumor types, since light scattering in deeper tumor complicates the analysis of the fluorescence signal. The T/B ratio calculated from shallow tumor types would better represent the differential expression of the biomarkers. However, in our previous study [135], when using T/B ratio to quantify IV injected 5-ALA for brain tumor resection, we achieved a mean (±SD) T/B ratio of 2.87±1.08, which is greater than the values reported in this dissertation. A direct comparison of the T/B ratio is difficult since the experiment setups are different, and different contrast agents are applied to different tumor types. Nonetheless, the higher T/B ratios were resulted from the compensation procedure before T/B calculation, which enhanced the target signal and suppressed the background noise. In the future, we will evaluate T/B ratio on IV injected contrast agents and other tumor types. Other factors, such as compensation method and background region selection (contralateral vs. surrounding), should also be considered for T/B calculation.

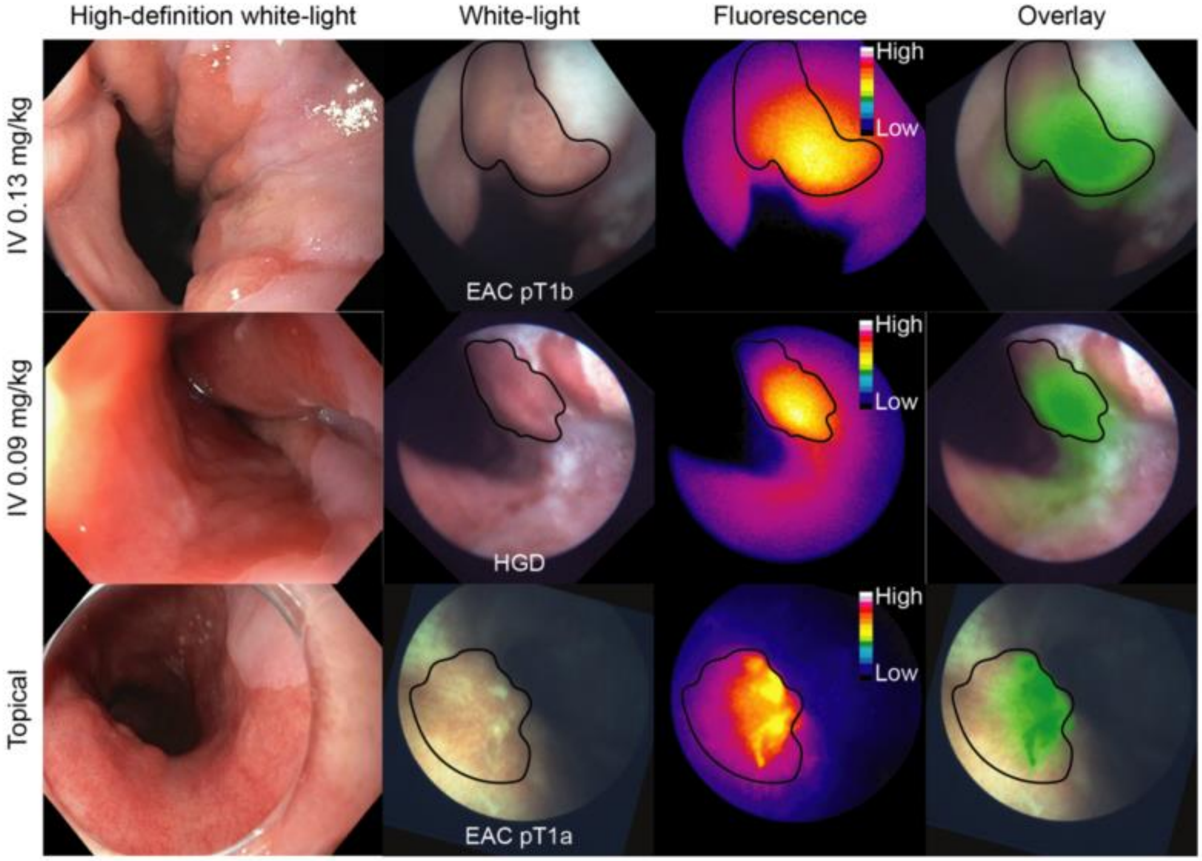


Figure 6.1 Representative images from de Jongh et al. study, C-Met targeted fluorescence contrast agents were either topically applied (bottom row) or IV injected (top 2 rows) for Barrett's neoplasia detection [134].

6.2.2 *Full color + NIR fluorescence mmSFE*

Concurrent fluorescence images with single channel reflectance images were captured for the studies in this dissertation. The reflectance images from 638nm have lower contrast compared to full color reflectance images, which limit the extraction of morphological features for further image analysis, e.g. object detection, registration between mmSFE and mother endoscope, 3D reconstruction etc. A full color + NIR fluorescence mmSFE provides both structural and functional imaging. The low tissue absorption, low tissue scattering, and low autofluorescence of NIR light allow imaging devices to see behind the tissue surface. In addition, there is no spectrum overlap between full color RGB and NIR fluorescence, which enables concurrent 4 channel imaging with minimum crosstalk. To provide the surgeon with intuitive and unbiased visualization of the target tissue, a robust scheme is needed for the real-time fusion of color images with the fluorescence signals. Figure 6.2 shows examples of different overlaid schemes for mmSFE, where transparent overlay and CLAHE are applied to increase the overlaid image contrast. Topics about good colormap selection is discussed in recent papers by Glatz et al. and Elliott et al. [136, 137]

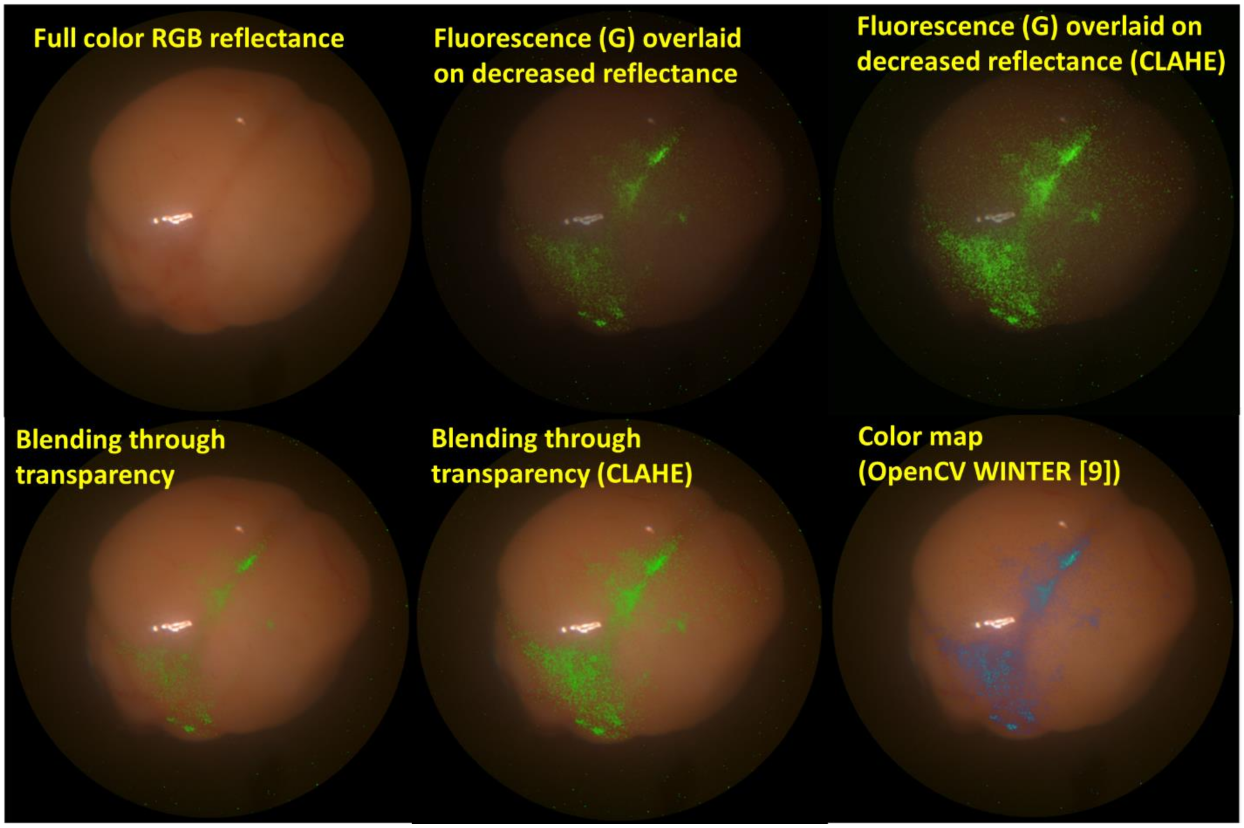


Figure 6.2 Examples of mmSFE fluorescence overlaid scheme

6.2.3 *Fluorescence calibration*

Molecular Fluorescence Endoscopy shows promising results in early cancer detection and guiding biopsy. Variable parameters and artifacts, such as ambient light, distance from the camera to the tissue, and specular reflection, diminish the robustness of quantitative analysis of the fluorescence image, e.g. T/B calculation. These variable parameters will also affect the performance of image processing, such as registration, 2D panorama image generation, 3D reconstruction. For example, the shadow generated by ununiform illumination will impair feature extraction, which is an essential for reconstruction.

Specular reflectance removal usually consists of two steps. First, the highlights are detected, either by threshold method or a more robust segmentation algorithm [138]. Second, the highlighted regions are recovered by interpolation, spatial [138] or temporal [139]. However, finding the corresponding position in preceding and subsequent frames as replacement is not always possible, especially when there is abrupt motion, or the light conditions change too much. Other novel techniques apply generative adversarial network to create paired image without specular reflection [140].

Distance compensation have been implemented to correct variations in illumination due to the variations in the distance and angle between the endoscope and the tissue (Chapter 3). The robustness of this ratiometric method is impaired by the tubular structure of the esophagus, where regions farther away along the esophagus appear darker in reflectance. The noise will be amplified when dividing fluorescence channel by reflectance. This error can be corrected by applying a mask to separate bright and dark regions before distance compensation.

6.2.4 *Object detection and instance segmentation*

In Chapter 5, a 2-step pipeline is applied for cancer classification. The frame selection and the fluorescence segmentation model are implemented sequentially for T/B calculation. However, only fluorescence intensity is considered when using T/B ratio for cancer classification, no morphological features are investigated, which may contain information for cancer progression. The discriminative features from white light endoscopy can be automatically learnt by deep learning techniques, to detection BE neoplasia. De Groof et al. developed a hybrid ResNet-UNet model CAD system for early BE neoplasia detection and localization. After classifying the images as neoplastic or not neoplastic, the CAD system then directly produces a heatmap and encircling the region suspicious for neoplasia to guide biopsy (Figure 6.3). To achieve this system with high performance, machine learning techniques, such as transfer learning and ensemble learning, were applied. The system detected neoplasia with high accuracy on 2 external validation data sets (data set 4, 80 patients: 89% accuracy, 90% sensitivity, and 88% specificity; data set 5, 80 patients: 88% accuracy, 93% sensitivity, and 83% specificity). The CAD system achieved higher classification accuracy than any of the individual 53 nonexpert endoscopists, with comparable delineation performance [141].

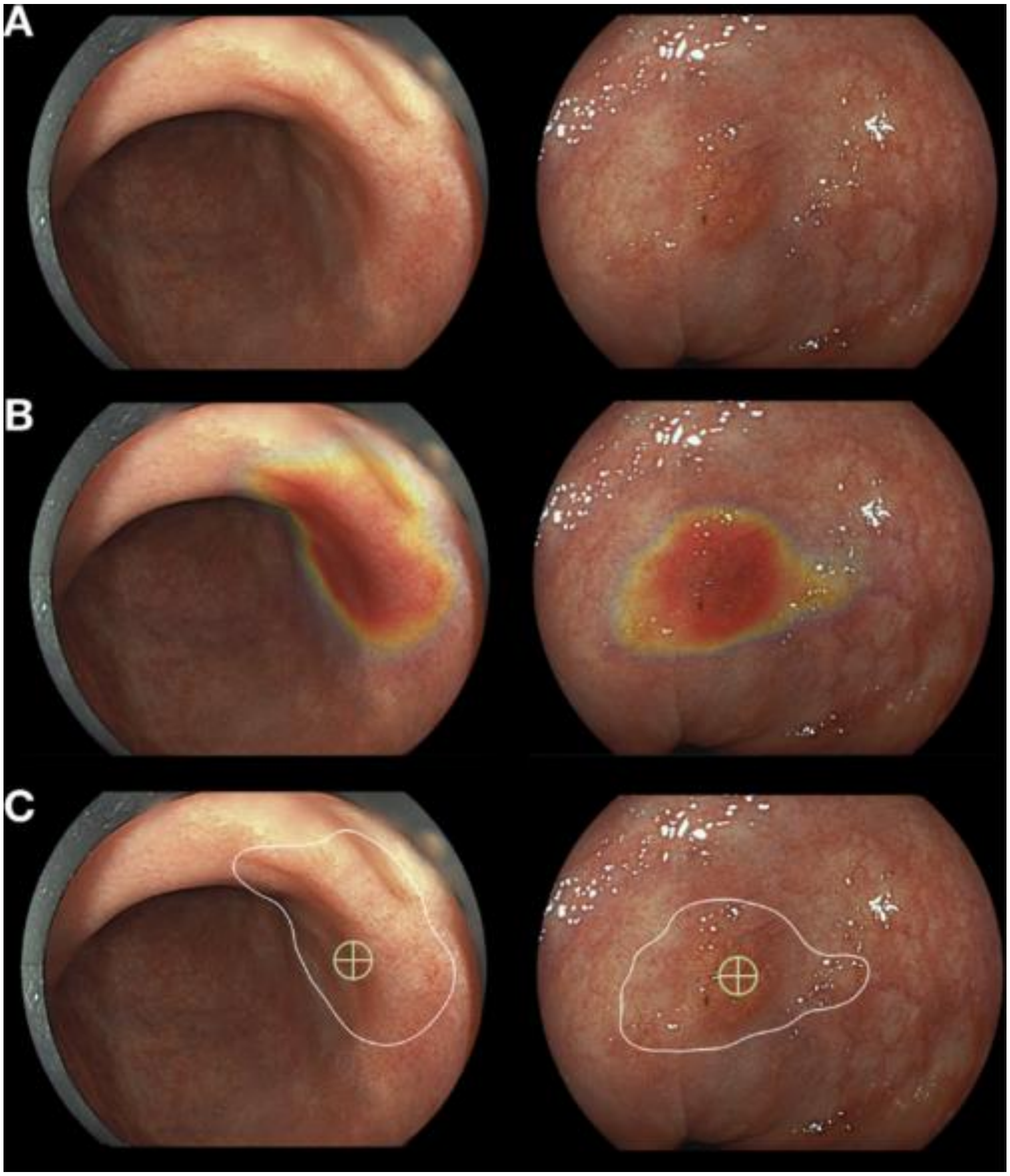


Figure 6.3 Example of (A) 2 neoplastic lesions with (B) the heatmap visualization by the CAD system and (C) its corresponding delineation and biopsy site indicator [141]

However, flat and subtle lesions could be missed by their CAD system. To improve the detection accuracy of the flat neoplastic lesions, targeted fluorophores can be applied in addition to WLE, which provide molecular contrast associated with cancer progression. The concurrent four channel mmSFE images (full color + NIR fluorescence) contain both structural and molecular information and can be used as input for deep learning CAD system. Architectures, such as Mask R-CNN

[142], can be used to locate and classify suspicious regions. Figure 6.4 gives a conceptual example how the Mask R-CNN algorithm will work for BE neoplasia classification and localization. The algorithm will localize each object in the image, including EAC/HGD lesions, dye pooling regions, forceps. Instance segmentation masks will be generated and overlaid on the WLE image. The fused image with neoplastic lesion boundary will be used for guiding biopsy. Additional T/B ratio can be calculated to better quantify the fluorescence signal. To provide real-time instance segmentation, algorithms, such as Yolact [143] and SOLOv2 [144], can be investigated. However, a large dataset is needed to train these networks for neoplasia classification and localization, which can be eventually fulfilled when the technique is utilized in a multi-center clinical trial for BE surveillance.

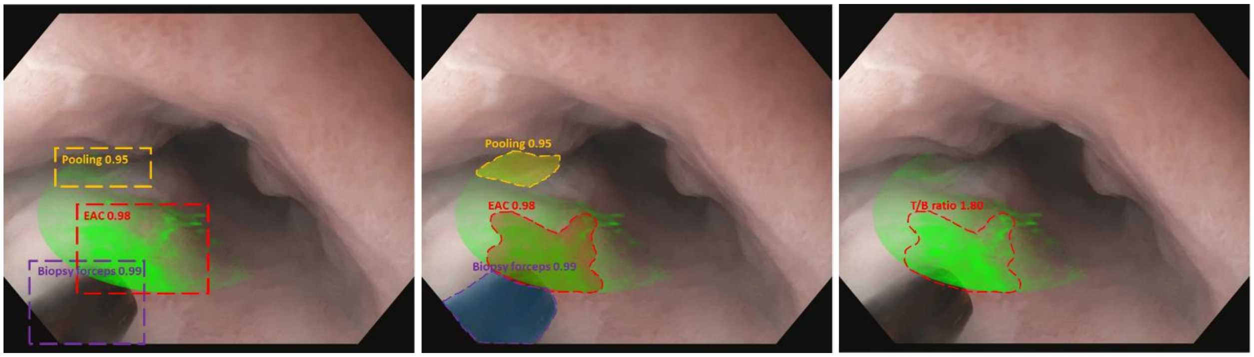


Figure 6.4 A conceptual example how the Mask R-CNN algorithm will work for cancer localization and classification.

6.2.5 *2D panorama and 3D reconstruction of multimodal endoscopic videos*

Multimodal endoscopic with automatic T/B calculation is suitable for routine cancer screening or surveillance in the esophagus. However, limited by the field of view and the restricted operating angle of the endoscope, the interior esophagus wall may not be completely examined. A 2D panorama or 3D reconstruction of endoscopic videos can be used to provide a larger field of view of esophagus. Once the 2D panorama or 3D esophagus with fluorescence targets is constructed from recorded video sequences, a gastroenterologist can then analyze remotely. Such reconstructed esophagus can also be used as time stamps for monitoring treatment and surveillance over time. When constructing the 2D panorama from a trans-nasal [145] or tethered capsule mmSFE [146], we expect a lower cost of the endoscopic procedure and a greater accessibility across high risk populations.

The 2D panorama or 3D reconstruction of esophagus is still challenging due to the lack of anatomical features. Methods have been applied in this field, including structure from motion (SfM) [147], simultaneous localization and mapping (SLAM), which requires texture features for registration. Other methods requiring hardware modifications are Structured Light [148] and Time-of-Flight [149]. Prinzen et al. proposed a two-step 3D reconstruction of the esophagus. A series of panorama images of the esophagus wall were generated using shape from shading approach, which are then aligned to create the 3D structure [150]. The overall pipeline for 3D multimodal esophagus reconstruction is shown in Figure 6.3.

Figure 6.5 Pipeline for 3D multimodal esophagus reconstruction

6.2.6 *New generation mmSFE*

With all the techniques developed and discussed in this dissertation, the new generation mmSFE should have features:

- Full color + NIR fluorescence concurrent imaging with robust overlaid scheme

- Real-time fluorescence calibration
- Real-time cancer classification and localization, with real-time automatic T/B calculation
- Real-time 2D panorama image of the esophagus with fluorescence overlaid
- Fluorescence-guided endoscopic biopsy ex vivo verification at the bedside [151]
- Offline feature to generate 3D reconstruction of the esophagus, which allows longitudinal treatment and surveillance monitor

Other potential features are super-resolution using deep learning [152], tumor margin delineation with threshold analysis [153].

BIBLIOGRAPHY

[1] American Cancer Society. Cancer Facts & Figures 2020.

[2] Vickers J. Role of endoscopic ultrasound in the preoperative assessment of patients with oesophageal cancer. *Ann R Coll Surg Engl*. 1998;80(4):233-239.

[3] Lehr L, Rupp N, Siewert JR. Assessment of resectability of esophageal cancer by computed tomography and magnetic resonance imaging. *Surgery*. 1988;103(3):344-350.

[4] Wiersema MJ, Wiersema LM. High-resolution 25-megahertz ultrasonography of the gastrointestinal wall: histologic correlates. *Gastrointest Endosc*. 1993;39(4):499-504. doi:10.1016/s0016-5107(93)70159-5

[5] Fass, L., 2008. Imaging and cancer: a review. Molecular oncology, 2(2), pp.115-152.

[6] Nguyen, Q.T. and Tsien, R.Y., 2013. Fluorescence-guided surgery with live molecular navigation—a new cutting edge. Nature reviews cancer, 13(9), p.653.

[7] Vahrmeijer, A.L., Hutteman, M., Van Der Vorst, J.R., Van De Velde, C.J. and Frangioni, J.V., 2013. Image-guided cancer surgery using near-infrared fluorescence. Nature reviews Clinical oncology, 10(9), p.507.

[8] Zhang, R.R., Schroeder, A.B., Grudzinski, J.J., Rosenthal, E.L., Warram, J.M., Pinchuk, A.N., Eliceiri, K.W., Kuo, J.S. and Weichert, J.P., 2017. Beyond the margins: real-time detection of cancer using targeted fluorophores. Nature Reviews Clinical Oncology, 14(6), p.347.

[9] McGirt, M.J., Chaichana, K.L., Attenello, F.J., Weingart, J.D., Than, K., Burger, P.C., Olivi, A., Brem, H. and Quinoñes-Hinojosa, A., 2008. Extent of surgical resection is independently associated with survival in patients with hemispheric infiltrating low-grade gliomas. Neurosurgery, 63(4), pp.700-708.

[10] Swindle, P., Eastham, J.A., Ohori, M., Kattan, M.W., Wheeler, T., Maru, N., Slawin, K. and Scardino, P.T., 2005. Do margins matter? The prognostic significance of positive surgical margins in radical prostatectomy specimens. The Journal of urology, 174(3), pp.903-907.

[11] https://clinicaltrials.gov/

[12] Cousins, A., Thompson, S.K., Wedding, A.B. and Thierry, B., 2014. Clinical relevance of novel imaging technologies for sentinel lymph node identification and staging. Biotechnology advances, 32(2), pp.269-279.

[13] Kobayashi, H. and Choyke, P.L., 2010. Target-cancer-cell-specific activatable fluorescence imaging probes: rational design and in vivo applications. Accounts of chemical research, 44(2), pp.83-90.

[14] Koch, M., Symvoulidis, P. and Ntziachristos, V., 2018. Tackling standardization in fluorescence molecular imaging. Nature Photonics, 12(9), pp.505-515.

[15] Tichauer, K.M., Samkoe, K.S., Sexton, K.J., Gunn, J.R., Hasan, T. and Pogue, B.W., 2012. Improved tumor contrast achieved by single time point dual-reporter fluorescence imaging. Journal of biomedical optics, 17(6), p.066001.

[16] Tichauer, K.M., Samkoe, K.S., Sexton, K.J., Hextrum, S.K., Yang, H.H., Klubben, W.S., Gunn, J.R., Hasan, T. and Pogue, B.W., 2012. In vivo quantification of tumor receptor binding potential with dual-reporter molecular imaging. Molecular Imaging and Biology, 14(5), pp.584-592.

[17] Howlader, N., Noone, A.M., Krapcho, M., Miller, D., Bishop, K., Altekruse, S.F., Kosary, C.L., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D.R., Chen, H.S., Feuer, E.J. and Cronin, K.A., “SEER Cancer Statistics Review, 1975-2013”, National Cancer Institute

[18] Hvid-Jensen, F., Pedersen, L., Drewes, A.M., Sørensen, H.T. and Funch-Jensen, P., “Incidence of adenocarcinoma among patients with Barrett's esophagus”, New England Journal of Medicine, 365(15), pp.1375-1383 (2011).

[19] Spechler, S. J., and R. K. Goyal, "The columnar-lined esophagus, intestinal metaplasia, and Norman Barrett," Gastroenterology 110(2), pp.614-621 (1996).

[20] Bhat, S., Coleman, H.G., Yousef, F., Johnston, B.T., McManus, D.T., Gavin, A.T. and Murray, L.J., “Risk of malignant progression in Barrett’s esophagus patients: results from a large population-based study,” Journal of the National Cancer Institute, 103(13), pp.1049-1057 (2011).

[21] Shaheen, N.J., Falk, G.W., Iyer, P.G. and Gerson, L.B., “ACG clinical guideline: diagnosis and management of Barrett’s esophagus,” The American Journal of Gastroenterology, 111(1), pp.30-50(2016).

[22] Abrams, J.A., Kapel, R.C., Lindberg, G.M., Saboorian, M.H., Genta, R.M., Neugut, A.I. and Lightdale, C.J., “Adherence to biopsy guidelines for Barrett's esophagus surveillance in the community setting in the United States,” Clinical Gastroenterology and Hepatology, 7(7), pp.736-742 (2009).

[23] Boerwinkel, D.F., Holz, J.A., Kara, M.A., Meijer, S.L., Wallace, M.B., Song, L.M.W.K., Ragunath, K., Wolfsen, H.C., Iyer, P.G., Wang, K.K. and Weusten, B.L., “Effects of autofluorescence imaging on detection and treatment of early neoplasia in patients with Barrett's esophagus,” Clinical Gastroenterology and Hepatology, 12(5), pp.774-781 (2014).

[24] Wolfsen, H.C., Crook, J.E., Krishna, M., Achem, S.R., Devault, K.R., Bouras, E.P., Loeb, D.S., Stark, M.E., Woodward, T.A., Hemminger, L.L. and Cayer, F.K., “Prospective, controlled tandem endoscopy study of narrow band imaging for dysplasia detection in Barrett's Esophagus,” Gastroenterology, 135(1), pp.24-31 (2008).

[25] Gora, M.J., Sauk, J.S., Carruth, R.W., Gallagher, K.A., Suter, M.J., Nishioka, N.S., Kava, L.E., Rosenberg, M., Bouma, B.E. and Tearney, G.J., “Tethered capsule endomicroscopy enables less invasive imaging of gastrointestinal tract microstructure,” Nature Medicine, 19(2), pp.238-240 (2013).

[26] Boerwinkel, D.F., Holz, J.A., Aalders, M.C.G., Visser, M., Meijer, S.L., Van Berge Henegouwen, M.I., Weusten, B.L.A.M. and Bergman, J.J.G.H.M., “Third-generation autofluorescence endoscopy for the detection of early neoplasia in Barrett's esophagus: a pilot study,” Diseases of the Esophagus, 27(3), pp.276-284 (2014).

[27] Muto, M., Katada, C., Sano, Y. and Yoshida, S., “Narrow band imaging: a new diagnostic approach to visualize angiogenesis in superficial neoplasia,” Clinical Gastroenterology and Hepatology, 3(7), pp.S16-S20 (2005).

[28] Curvers, W.L., Singh, R., Song, L.W.K., Wolfsen, H.C., Ragunath, K., Wang, K., Wallace, M.B., Fockens, P. and Bergman, J.J.G.H.M., “Endoscopic tri-modal imaging for detection of early neoplasia in Barrett’s oesophagus: a multi-centre feasibility study using high-resolution endoscopy, autofluorescence imaging and narrow band imaging incorporated in one endoscopy system,” Gut, 57(2), pp.167-172 (2008).

[29] Terry, N.G., Zhu, Y., Rinehart, M.T., Brown, W.J., Gebhart, S.C., Bright, S., Carretta, E., Ziefle, C.G., Panjehpour, M., Galanko, J. and Madanick, R.D., “Detection of dysplasia in Barrett's esophagus with in vivo depth-resolved nuclear morphology measurements,” Gastroenterology, 140(1), pp.42-50 (2011).

[30] Wolfsen, H.C., Sharma, P., Wallace, M.B., Leggett, C., Tearney, G. and Wang, K.K., “Safety and feasibility of volumetric laser endomicroscopy in patients with Barrett’s esophagus (with videos),” Gastrointestinal Endoscopy, 82(4), pp.631-640 (2015).

[31] Sturm, M.B., Joshi, B.P., Lu, S., Piraka, C., Khondee, S., Elmunzer, B.J., Kwon, R.S., Beer, D.G., Appelman, H.D., Turgeon, D.K. and Wang, T.D., “Targeted imaging of esophageal neoplasia with a fluorescently labeled peptide: first-in-human results,” Science Translational Medicine, 5(184), pp.184ra61-184ra61 (2013).

[32] Hsiung, P.L., Hardy, J., Friedland, S., Soetikno, R., Du, C.B., Wu, A.P., Sahbaie, P., Crawford, J.M., Lowe, A.W., Contag, C.H. and Wang, T.D., “Detection of colonic dysplasia in vivo using a targeted heptapeptide and confocal microendoscopy,” Nature Medicine, 14(4), pp.454-458 (2008).

[33] Leggett, C.L., Gorospe, E.C., Chan, D.K., Muppa, P., Owens, V., Smyrk, T.C., Anderson, M., Lutzke, L.S., Tearney, G. and Wang, K.K., “Comparative diagnostic performance of volumetric laser endomicroscopy and confocal laser endomicroscopy in the

detection of dysplasia associated with Barrett’s esophagus,” Gastrointestinal Endoscopy, 83(5), pp.880-888 (2016).

[34] Bird-Lieberman, E.L., Neves, A.A., Lao-Sirieix, P., O'Donovan, M., Novelli, M., Lovat, L.B., Eng, W.S., Mahal, L.K., Brindle, K.M. and Fitzgerald, R.C., “Molecular imaging using fluorescent lectins permits rapid endoscopic identification of dysplasia in Barrett's esophagus,” Nature Medicine, 18(2), pp.315-321 (2012).

[35] Joshi, B.P., Duan, X., Kwon, R.S., Piraka, C., Elmunzer, B.J., Lu, S., Rabinsky, E.F., Beer, D.G., Appelman, H.D., Owens, S.R. and Kuick, R., “Multimodal endoscope can quantify wide-field fluorescence detection of Barrett’s neoplasia,” Endoscopy, 48(02), pp.A1-A13 (2016).

[36] Qi, X., Sivak, M.V., Isenberg, G., Willis, J.E. and Rollins, A.M., “Computer-aided diagnosis of dysplasia in Barrett’s esophagus using endoscopic optical coherence tomography,” Journal of Biomedical Optics, 11(4), pp.044010-044010 (2006).

[37] Chan, P.K., Cheng, S.H. and Poon, T.C., “Automated segmentation in confocal images using a density clustering method,” Journal of Electronic Imaging, 16(4), pp.043003-043003 (2007).

[38] Chan, T.F. and Vese, L.A., “Active contours without edges,” IEEE Transactions on Image Processing, 10(2), pp.266-277 (2001).

[39] Hajian-Tilaki, K., “Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation”, Caspian Journal of Internal Medicine, 4(2), p.627 (2013).

[40] Sharma, P., Savides, T.J., Canto, M.I., Corley, D.A., Falk, G.W., Goldblum, J.R., Wang, K.K., Wallace, M.B. and Wolfsen, H.C., “The American Society for Gastrointestinal Endoscopy PIVI (preservation and incorporation of valuable endoscopic innovations) on imaging in Barrett's esophagus”, Gastrointestinal Endoscopy, 76(2), pp.252-254 (2012).

[41] Lee, C.M., Engelbrecht, C.J., Soper, T.D., Helmchen, F. and Seibel, E.J., “Scanning fiber endoscopy with highly flexible, 1-mm catheterscopes for wide-field, full-color imaging”, Journal of Biophotonics, 3(5-6), p.385 (2010).

[42] Yang, C., Hou, V., Nelson, L.Y. and Seibel, E.J., “Color-matched and fluorescence-labeled esophagus phantom and its applications”, Journal of Biomedical Optics, 18(2), pp.026020-026020 (2013)

[43] Yang, C., Hou, V., Nelson, L.Y. and Seibel, E.J., “Mitigating fluorescence spectral overlap in wide-field endoscopic imaging”, Journal of Biomedical Optics, 18(8), pp.086012-086012 (2013).

[44] Yang, C., Hou, V.W., Girard, E.J., Nelson, L.Y. and Seibel, E.J., “Target-to-background enhancement in multispectral endoscopy with background autofluorescence mitigation for quantitative molecular imaging”, Journal of biomedical optics, 19(7), pp.076014-076014 (2014).

[45] Joshi, B. P., Miller, S. J., Lee, C. M., Seibel, E. J., and Wang, T. D., “Multispectral endoscopic imaging of colorectal dysplasia in vivo”, Gastroenterology, 143(6), 1435–1437 (2012).

[46] Lacroix, M., Abi-Said, D., Fourney, D.R., Gokaslan, Z.L., Shi, W., DeMonte, F., Lang, F.F., McCutcheon, I.E., Hassenbusch, S.J., Holland, E. and Hess, K., 2001. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. Journal of neurosurgery, 95(2), pp.190-198.

[47] McGirt, M.J., Chaichana, K.L., Gathinji, M., Attenello, F.J., Than, K., Olivi, A., Weingart, J.D., Brem, H. and Quiñones-Hinojosa, A.R., 2009. Independent association of extent of resection with survival in patients with malignant brain astrocytoma. Journal of neurosurgery, 110(1), pp.156-162.

[48] Stummer, W., Pichlmeier, U., Meinel, T., Wiestler, O.D., Zanella, F., Reulen, H.J. and ALA-Glioma Study Group, 2006. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. The lancet oncology, 7(5), pp.392-401.

[49] Samkoe, K.S., Bates, B.D., Elliott, J.T., LaRochelle, E., Gunn, J.R., Marra, K., Feldwisch, J., Ramkumar, D.B., Bauer, D.F., Paulsen, K.D. and Pogue, B.W., 2018. Application of fluorescence-guided surgery to subsurface cancers requiring wide local excision: literature review and novel developments toward indirect visualization. Cancer Control, 25(1), p.1073274817752332.

[50] Grotz, T.E., Markovic, S.N., Erickson, L.A., Harmsen, W.S., Huebner, M., Farley, D.R., Pockaj, B.A., Donohue, J.H., Sim, F.H., Grant, C.S. and Bagaria, S.P., 2011, June. Mayo Clinic consensus recommendations for the depth of excision in primary cutaneous melanoma. In Mayo Clinic Proceedings (Vol. 86, No. 6, pp. 522-528). Elsevier.

[51] Martirosyan, N.L., Cavalcanti, D.D., Eschbacher, J.M., Delaney, P.M., Scheck, A.C., Abdelwahab, M.G., Nakaji, P., Spetzler, R.F. and Preul, M.C., 2011. Use of in vivo near-infrared laser confocal endomicroscopy with indocyanine green to detect the boundary of infiltrative tumor. Journal of neurosurgery, 115(6), pp.1131-1138.

[52] Hamdoon, Z., Jerjes, W., McKenzie, G., Jay, A. and Hopper, C., 2016. Optical coherence tomography in the assessment of oral squamous cell carcinoma resection margins. Photodiagnosis and photodynamic therapy, 13, pp.211-217.

[53] Fabelo, H., Ortega, S., Lazcano, R., Madroñal, D., M Callicó, G., Juárez, E., Salvador, R., Bulters, D., Bulstrode, H., Szolna, A. and Piñeiro, J.F., 2018. An intraoperative visualization system using hyperspectral imaging to aid in brain tumor delineation. Sensors, 18(2), p.430.

[54] Jiang, L., Wang, X., Wu, Z., Du, H., Wang, S., Li, L., Fang, N., Lin, P., Chen, J., Kang, D. and Zhuo, S., 2017. Label-free imaging of brain and brain tumor specimens with combined two-photon excited fluorescence and second harmonic generation microscopy. Laser Physics Letters, 14(10), p.105401.

[55] Karabeber, H., Huang, R., Iacono, P., Samii, J.M., Pitter, K., Holland, E.C. and Kircher, M.F., 2014. Guiding brain tumor resection using surface-enhanced Raman scattering nanoparticles and a hand-held Raman scanner. Acs Nano, 8(10), pp.9755-9766.

[56] Jiang, Y., Girard, E.J., Pakiam, F. and Seibel, E.J., 2018, March. Ultrathin and flexible 4-channel scope for guiding surgical resections using a near-infrared fluorescence molecular probe for cancer. In Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 10576, p. 105762K). International Society for Optics and Photonics.

[57] Hartmans, E., Tjalma, J.J., Linssen, M.D., Allende, P.B.G., Koller, M. and Jorritsma-Smit, A., 2018. Potential red-flag identification of colorectal adenomas with wide-field fluorescence molecular endoscopy. Theranostics, 8(6), p.1458.

[58] Erickson-Bhatt, S.J., Nolan, R.M., Shemonski, N.D., Adie, S.G., Putney, J., Darga, D., McCormick, D.T., Cittadine, A.J., Zysk, A.M., Marjanovic, M. and Chaney, E.J., 2015. Real-time imaging of the resection bed using a handheld probe to reduce incidence of microscopic positive margins in cancer surgery. Cancer research, 75(18), pp.3706-3712.

[59] Acerbi, F., Broggi, M., Eoli, M., Anghileri, E., Cuppini, L., Pollo, B., Schiariti, M., Visintini, S., Orsi, C., Franzini, A. and Broggi, G., 2013. Fluorescein-guided surgery for grade IV gliomas with a dedicated filter on the surgical microscope: preliminary results in 12 cases. Acta neurochirurgica, 155(7), pp.1277-1286.

[60] Wang, X., Bhaumik, S., Staudinger, V.P. and Yazdanfar, S., 2010. Compact instrument for fluorescence image-guided surgery. Journal of biomedical optics, 15(2), p.020509.

[61] Lee, J.Y., Pierce, J.T., Thawani, J.P., Zeh, R., Nie, S., Martinez-Lage, M. and Singhal, S., “Near-infrared fluorescent image-guided surgery for intracranial meningioma,” Journal of Neurosurgery, pp.1-11 (2017).

[62] Hendricks, B.K., Sanai, N. and Stummer, W., 2018. Fluorescence-guided surgery with aminolevulinic acid for low-grade gliomas. Journal of neuro-oncology, pp.1-6.

[63] Stummer, W., Novotny, A., Stepp, H., Goetz, C., Bise, K. and Reulen, H.J., 2000. Fluorescence-guided resection of glioblastoma multiforme utilizing 5-ALA-induced porphyrins: a prospective study in 52 consecutive patients. Journal of neurosurgery, 93(6), pp.1003-1013.

[64] Belykh, E., Miller, E.J., Hu, D., Martirosyan, N.L., Woolf, E.C., Scheck, A.C., Byvaltsev, V.A., Nakaji, P., Nelson, L.Y., Seibel, E.J. and Preul, M.C., 2018. Scanning Fiber Endoscope Improves Detection of 5-Aminolevulinic Acid–Induced Protoporphyrin IX Fluorescence at the Boundary of Infiltrative Glioma. World neurosurgery, 113, pp.e51-e69.

[65] Rosenthal, E.L., Warram, J.M., de Boer, E., Chung, T.K., Korb, M.L., Brandwein-Gensler, M., Strong, T.V., Schmalbach, C.E., Morlandt, A.B., Agarwal, G. and Hartman, Y.E., “Safety and tumor specificity of cetuximab-IRDye800 for surgical navigation in head and neck cancer,” Clinical Cancer Research, 21(16), pp.3658-3666 (2015).

[66] de Souza, A.L.R., Marra, K., Gunn, J., Samkoe, K.S., Hoopes, P.J., Feldwisch, J., Paulsen, K.D. and Pogue, B.W., “Fluorescent Affibody molecule administered in vivo at a microdose level labels EGFR expressing glioma tumor regions,” Molecular Imaging and Biology, 19(1), pp.41-48 (2017).

[67] Butte, P.V., Mamelak, A., Parrish-Novak, J., Drazin, D., Shweikeh, F., Gangalum, P.R., Chesnokova, A., Ljubimova, J.Y. and Black, K., “Near-infrared imaging of brain tumors using the Tumor Paint BLZ-100 to achieve near-complete resection of brain tumors,” Neurosurgical focus, 36(2), p.E1 (2014).

[68] Fidel, J., Kennedy, K.C., Dernell, W.S., Hansen, S., Wiss, V., Stroud, M.R., Molho, J.I., Knoblaugh, S.E., Meganck, J., Olson, J.M. and Rice, B., “Preclinical validation of the utility of BLZ-100 in providing fluorescence contrast for imaging spontaneous solid tumors.,” Cancer research, 75(20), pp.4283-4291 (2015).

[69] Elliott, J.T., Marra, K., Evans, L.T., Davis, S.C., Samkoe, K.S., Feldwisch, J., Paulsen, K.D., Roberts, D.W. and Pogue, B.W., 2016. Simultaneous in vivo fluorescent markers for perfusion, protoporphyrin metabolism, and EGFR expression for optically guided identification of orthotopic glioma. Clinical Cancer Research.

[70] Lu, G., Halig, L., Wang, D., Chen, Z.G. and Fei, B., 2014, March. Hyperspectral imaging for cancer surgical margin delineation: registration of hyperspectral and histological images. In Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 9036, p. 90360S). International Society for Optics and Photonics.

[71] Brabetz, S., Leary, S.E., Gröbner, S.N., Nakamoto, M.W., Şeker-Cin, H., Girard, E.J., Cole, B., Strand, A.D., Bloom, K.L., Hovestadt, V. and Mack, N.L., 2018. A biobank of patient-derived pediatric brain tumor models. Nature Medicine, p.1.

[72] Parrish-Novak, J., Byrnes-Blake, K., Lalayeva, N., Burleson, S., Fidel, J., Gilmore, R., Gayheart-Walsten, P., Bricker, G.A., Crumb Jr, W.J., Tarlo, K.S. and Hansen, S., 2017. Nonclinical profile of BLZ-100, a tumor-targeting fluorescent imaging agent. International journal of toxicology, 36(2), pp.104-112.

[73] Johns, M., Giller, C.A., German, D.C. and Liu, H., “Determination of reduced scattering coefficient of biological tissue from a needle-like probe,” Opt. Express 13(13), 4828–4842 (2005).

[74] Abdo, A., Ersen, A. and Sahin, M., “Near-infrared light penetration profile in the rodent brain,” Journal of Biomedical Optics, 18(7), 075001 (2013).

[75] Jiang, Y., Gong, Y., Rubenstein, J.H., Wang, T.D. and Seibel, E.J., 2017. Toward real-time quantification of fluorescence molecular probes using target/background ratio for guiding biopsy and endoscopic therapy of esophageal neoplasia. Journal of Medical Imaging, 4(2), p.024502.

[76] Lee, C.M., Engelbrecht, C.J., Soper, T.D., Helmchen, F. and Seibel, E.J., “Scanning fiber endoscopy with highly flexible, 1 mm catheterscopes for wide-field, full-color imaging,” Journal of biophotonics, 3(5-6), pp.385-407 (2010).

[77] Barolet, D., 2008. Light-emitting diodes (LEDs) in dermatology. , 27, 4, 27(4), pp.227-238.

[78] Welch, A.J., Gardner, C., Richards‐Kortum, R., Chan, E., Criswell, G., Pfefer, J. and Warren, S., 1997. Propagation of fluorescent light. Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery, 21(2), pp.166-178.

[79] Unger, J., Sun, T., Chen, Y.L., Phipps, J.E., Bold, R.J., Darrow, M.A., Ma, K.L. and Marcu, L., 2018. Method for accurate registration of tissue autofluorescence imaging data with corresponding histology: a means for enhanced tumor margin assessment. Journal of biomedical optics, 23(1), p.015001.

[80] Wang, C.W., Ka, S.M. and Chen, A., 2014. Robust image registration of biological microscopic images. Scientific reports, 4, p.6050.

[81] Lurie, K.L., Gurjarpadhye, A.A., Seibel, E.J. and Ellerbee, A.K., 2015. Rapid scanning catheterscope for expanded forward-view volumetric imaging with optical coherence tomography. Optics letters, 40(13), pp.3165-3168.

[82] Feroldi, F., Verlaan, M., Knaus, H., Davidoiu, V., Vugts, D.J., van Dongen, G.A., Molthoff, C.F. and de Boer, J.F., 2018. High resolution combined molecular and structural optical imaging of colorectal cancer in a xenograft mouse model. Biomedical Optics Express, 9(12), pp.6186-6204.

[83] Zhou, J., Joshi, B.P., Duan, X., Pant, A., Qiu, Z., Kuick, R., Owens, S.R. and Wang, T.D., 2015. EGFR overexpressed in colonic neoplasia can be detected on wide-field endoscopic imaging. Clinical and translational gastroenterology, 6(7), p.e101.

[84] Miller, S.J., Lee, C.M., Joshi, B.P., Gaustad, A., Seibel, E.J. and Wang, T.D., 2012. Special Section on Endomicroscopy Technologies and Biomedical Applications: Targeted detection of murine colonic dysplasia in vivo with flexible multispectral scanning fiber endoscopy. Journal of Biomedical Optics, 17(2).

[85] Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology, 26(3), pp.297-302.

[86] Pogue, B.W., Rosenthal, E.L., Achilefu, S. and van Dam, G.M., 2018. Perspective review of what is needed for molecular-specific fluorescence-guided surgery. Journal of biomedical optics, 23(10), p.100601.

[87] Goshtasby, A., 1986. Piecewise linear mapping functions for image registration. Pattern Recognition, 19(6), pp.459-466.

[88] Vogelstein, B. et al. Cancer genome landscapes. Science 339, 1546-1558 (2013).

[89] Valdes, P.A., Leblond, F., Paulsen, K.D., Kim, A., Wilson, B.C., Conde, O.M., Harris, B.T. and Roberts, D.W., 2011. Combined fluorescence and reflectance spectroscopy for in vivo quantification of cancer biomarkers in low-and high-grade glioma surgery. Journal of biomedical optics, 16(11), p.116007.

[90] Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin Med Res. 7(1/2):4-13. PMID: 19574486.

[91] Citri, A. & Yarden, Y. EGF-ERBB signalling: towards the systems level. Nat. Rev. Mol. Cell. Biol. 7, 505-516 (2006).

[92] Dulak, A. M. et al. Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. Cancer Res. 72, 4383-4393 (2012).

[93] Dulak, A. M. et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nat. Genet. 45, 478-486 (2013).

[94] Miller, C. T. et al. Gene amplification in esophageal adenocarcinomas and Barrett's with high-grade dysplasia. Clin. Cancer Res. 9, 4819-4825 (2003)

[95] Joshi, B. P. et al. Design and synthesis of near-infrared peptide for in vivo molecular imaging of HER2. Bioconjug. Chem. 27, 481-494 (2016).

[96] Jiang, Y., Gong, Y., Rubenstein, J. H., Wang, T. D. & Seibel, E. J. Toward real-time quantification of fluorescence molecular probes using target/background ratio for guiding biopsy and endoscopic therapy of esophageal neoplasia. J. Med. Imaging 4, 024502 (2017).

[97] Bradski, G. & Kaehler, A. in learning OpenCV (O'Reilly media, 2008).

[98] Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI. 14, 1137-1145 (1995).

[99] Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J. Intern. Med. 4, 627 (2013).

[100] Pedregosa, F. et al. Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12, 2825-2830 (2011).

[101] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[102] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[103] Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

[104] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. and Bengio, Y., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[105] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. and Dieleman, S., 2016. Mastering the game of Go with deep neural networks and tree search. nature, 529(7587), p.484.

[106] Min, S., Lee, B. and Yoon, S., 2017. Deep learning in bioinformatics. Briefings in bioinformatics, 18(5), pp.851-869.

[107] Milletari, F., Ahmadi, S.A., Kroll, C., Plate, A., Rozanski, V., Maiostre, J., Levin, J., Dietrich, O., Ertl-Wagner, B., Bötzel, K. and Navab, N., 2017. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. Computer Vision and Image Understanding, 164, pp.92-102.

[108] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B. and Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Medical image analysis, 42, pp.60-88.

[109] Litjens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B. and Van Der Laak, J., 2016. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Scientific reports, 6, p.26286.

[110] Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G. and Naidich, D.P., 2019. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature medicine, 25(6), pp.954-961.

[111] Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A. and Karssemeijer, N., 2017. Large scale deep learning for computer aided detection of mammographic lesions. Medical image analysis, 35, pp.303-312.

[112] Trebeschi, S., van Griethuysen, J.J., Lambregts, D.M., Lahaye, M.J., Parmar, C., Bakers, F.C., Peters, N.H., Beets-Tan, R.G. and Aerts, H.J., 2017. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. Scientific reports, 7(1), pp.1-9.

[113] Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T. and Keutzer, K., 2014. Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869.

[114] Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[115] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W. and Heng, P.A., 2018. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE transactions on medical imaging, 37(12), pp.2663-2674.

[116] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M. and Larochelle, H., 2017. Brain tumor segmentation with deep neural networks. Medical image analysis, 35, pp.18-31.

[117] Wang, P., Xiao, X., Brown, J.R.G., Berzin, T.M., Tu, M., Xiong, F., Hu, X., Liu, P., Song, Y., Zhang, D. and Yang, X., 2018. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. Nature biomedical engineering, 2(10), pp.741-748.

[118] Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W. and Baldi, P., 2018. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology, 155(4), pp.1069-1078.

[119] Ali, S., Zhou, F., Braden, B., Bailey, A., Yang, S., Cheng, G., Zhang, P., Li, X., Kayser, M., Soberanis-Mukul, R.D. and Albarqouni, S., 2020. An objective comparison of detection and segmentation algorithms for artefacts in clinical endoscopy. Scientific reports, 10(1), pp.1-15.

[120] Izadyyazdanabadi, M., Belykh, E., Mooney, M.A., Eschbacher, J.M., Nakaji, P., Yang, Y. and Preul, M.C., 2018. Prospects for theranostics in neurosurgical imaging: empowering confocal laser endomicroscopy diagnostics via deep learning. Frontiers in oncology, 8, p.240.

[121] Caicedo, J.C., Roth, J., Goodman, A., Becker, T., Karhohs, K.W., Broisin, M., Molnar, C., McQuin, C., Singh, S., Theis, F.J. and Carpenter, A.E., 2019. Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. Cytometry Part A, 95(9), pp.952-965.

[122] Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M. and Van Valen, D., 2019. Deep learning for cellular image analysis. Nature methods, pp.1-14.

[123] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G. and Sang, N., 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 325-341).

[124] Carles, Guillem, Paul Zammit, and Andrew R. Harvey. "Holistic monte-carlo optical modelling of biological imaging." Scientific reports 9, no. 1 (2019): 1-11.

[125] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167 (2015).

[126] Ramachandran, P., Zoph, B. and Le, Q.V., 2017. Swish: a self-gated activation function. arXiv preprint arXiv:1710.05941, 7.

[127] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[128] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[129] Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

[130] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

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

[132] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." In Advances in neural information processing systems, pp. 2672-2680. 2014.

[133] Shin, Hoo-Chang, Neil A. Tenenholtz, Jameson K. Rogers, Christopher G. Schwarz, Matthew L. Senjem, Jeffrey L. Gunter, Katherine P. Andriole, and Mark Michalski. "Medical image synthesis for data augmentation and anonymization using generative adversarial networks." In International workshop on simulation and synthesis in medical imaging, pp. 1-11. Springer, Cham, 2018.

[134] de Jongh, Steven J., Floris J. Voskuil, Iris Schmidt, Arend Karrenbeld, Gursah Kats-Ugurlu, Gert Jan Meersma, Jessie Westerhof et al. "C-Met targeted fluorescence molecular endoscopy in Barrett's esophagus patients and identification of outcome parameters for phase-I studies." Theranostics 10, no. 12 (2020): 5357.

[135] Hu, Danying, Yang Jiang, Evgenii Belykh, Yuanzheng Gong, Mark C. Preul, Blake Hannaford, and Eric J. Seibel. "Toward real-time tumor margin identification in image-guided robotic brain tumor resection." In Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 10135, p. 101350D. International Society for Optics and Photonics, 2017.

[136] Glatz, J., Symvoulidis, P., Garcia-Allende, P.B. and Ntziachristos, V., 2014. Robust overlay schemes for the fusion of fluorescence and color channels in biological imaging. Journal of biomedical optics, 19(4), p.040501.F

[137] Elliott, J.T., Dsouza, A.V., Davis, S.C., Olson, J.D., Paulsen, K.D., Roberts, D.W. and Pogue, B.W., 2015. Review of fluorescence guided surgery visualization and overlay techniques. Biomedical optics express, 6(10), pp.3765-3782.

[138] Arnold, M., Ghosh, A., Ameling, S. and Lacey, G., 2010. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. EURASIP Journal on Image and Video Processing, 2010, pp.1-12.

[139] Stoyanov, D. and Yang, G.Z., 2005, September. Removing specular reflection components for robotic assisted laparoscopic surgery. In IEEE International Conference on Image Processing 2005 (Vol. 3, pp. III-632). IEEE.

[140] Funke, I., Bodenstedt, S., Riediger, C., Weitz, J. and Speidel, S., 2018, March. Generative adversarial networks for specular highlight removal in endoscopic images. In Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling (Vol. 10576, p. 1057604). International Society for Optics and Photonics.

[141] de Groof, A.J., Struyvenberg, M.R., van der Putten, J., van der Sommen, F., Fockens, K.N., Curvers, W.L., Zinger, S., Pouw, R.E., Coron, E., Baldaque-Silva, F. and Pech, O., 2020. Deep-learning system detects neoplasia in patients with barrett’s esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. Gastroenterology, 158(4), pp.915-929.

[142] He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn." In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969. 2017.

[143] Bolya, Daniel, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. "Yolact: Real-time instance segmentation." In Proceedings of the IEEE international conference on computer vision, pp. 9157-9166. 2019.

[144] Wang, Xinlong, Rufeng Zhang, Tao Kong, Lei Li, and Chunhua Shen. "SOLOv2: Dynamic, Faster and Stronger." arXiv preprint arXiv:2003.10152 (2020).

[145] Peery, Anne F., Toshitaka Hoppo, Katherine S. Garman, Evan S. Dellon, Norma Daugherty, Susan Bream, Alejandro F. Sanz et al. "Feasibility, safety, acceptability, and yield of office-based, screening transnasal esophagoscopy (with video)." Gastrointestinal endoscopy 75, no. 5 (2012): 945-953.

[146] Seibel, Eric J., Robert E. Carroll, Jason A. Dominitz, Richard S. Johnston, C. David Melville, Cameron M. Lee, Steven M. Seitz, and Michael B. Kimmey. "Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and Barrett's esophagus." IEEE Transactions on Biomedical Engineering 55, no. 3 (2008): 1032-1042.

[147] Lurie, K.L., Angst, R., Zlatev, D.V., Liao, J.C. and Bowden, A.K.E., 2017. 3D reconstruction of cystoscopy videos for comprehensive bladder records. Biomedical optics express, 8(4), pp.2106-2123.

[148] Ackerman, J.D., Keller, K. and Fuchs, H., 2002, March. Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality. In Three-Dimensional Image Capture and Applications V (Vol. 4661, pp. 39-46). International Society for Optics and Photonics.

[149] Penne, J., Höller, K., Stürmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feußner, H., Schmauss, B. and Hornegger, J., 2009, September. Time-of-flight 3-D endoscopy. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 467-474). Springer, Berlin, Heidelberg.

[150] Prinzen, M., Trost, J., Bergen, T., Nowack, S. and Wittenberg, T., 2015. 3D Shape Reconstruction of the Esophagus from Gastroscopic Video. In Bildverarbeitung für die Medizin 2015 (pp. 173-178). Springer Vieweg, Berlin, Heidelberg.

[151] Fauver et al. " Design and testing of fluorescence-guided endoscopic biopsy assessment at the bedside." In Design and Quality for Biomedical Technologies XIV, under review, International Society for Optics and Photonics, 2021.

[152] Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., Bentolila, L.A., Kural, C. and Ozcan, A., 2019. Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature Methods, 16(1), pp.103-110.

[153] Koch, M., de Jong, J.S., Glatz, J., Symvoulidis, P., Lamberts, L.E., Adams, A.L., Kranendonk, M.E., van Scheltinga, A.G.T., Aichler, M., Jansen, L. and de Vries, J., 2017. Threshold analysis and biodistribution of fluorescently labeled bevacizumab in human breast cancer. Cancer research, 77(3), pp.623-631.