Towards Automatic Identification and Delineation of Tissues and Pathologies in H&E Stained Images

Ramamurthy Bhagavatula

Advisor: Prof. Jelena Kovačević

Department of Electrical and Computer Engineering
Carnegie Institute of Technology
Carnegie Mellon University, Pittsburgh, PA 15213

September, 2011

Thesis Manuscript

Submitted in partial fulfillment of the requirements towards the Ph.D. degree awarded by the Department of Electrical and Computer Engineering, Carnegie Inst. of Tech., Carnegie Mellon University.

Thesis Committee Members

Jelena Kovačević (Advisor)
  Departments of Biomedical Engineering and
  Electrical and Computer Engineering,
  Carnegie Mellon University

Matthew C. Fickus
  Department of Mathematics and Statistics,
  Air Force Institute of Technology

Markus Püschel
  Department of Computer Science,
  ETH Zurich, Switzerland

Gustavo K. Rohde
  Departments of Biomedical Engineering and
  Electrical and Computer Engineering,
  Carnegie Mellon University

Abstract

We propose a framework and methodology for the automated identification and delineation of tissues and their pathologies in hematoxylin and eosin (H&E) stained images. Histopathology is vital to medicine and research as it enables quantitative and qualitative analysis of tissue samples, stained and visualized via microscopes; the most routine and cost-effective of these stains is H&E. In clinical diagnostic surgical pathology, the pathologist interprets H&E-stained tissue slides by determining whether a given sample represents normal or abnormal tissue for the given anatomical location. Although pathologists accurately and consistently identify and delineate such tissues and their pathologies, this is a time-consuming and expensive task; thus the need for automated algorithms for improved throughput and robustness.

We develop such an algorithm that uses local histograms and occlusion models as a mathematical framework for pixel-level classification. We also develop an expert-guided feature set called the histopathology vocabulary that mimics the visual process used by pathologists. To expand applicability, we achieve simultaneous identification and delineation by performing pixel-level classification.

Experimental results on both a clinical application (active colitis) and a research one (tissue development in teratoma tumors) validate the discriminative power of our approach. We also present comparisons to popular, though general, feature types to demonstrate the power of our expert-guided feature set. Our framework and methodology demonstrates great promise towards the creation of a framework and methodology for the automated

identification and delineation of tissues and their pathologies in H&E-stained images.

Acknowledgments

I want to express my utmost appreciation to Jelena Kovačević for being the best advisor I could have hoped for and a good friend. When I asked to join her group, she welcomed me with open arms and a level of excitement that would prove to be infectious. Whether it was in her courses or in the pursuit of this thesis, her invaluable lessons and advice constantly kept me moving forward. The countless interesting projects she let me participate in made my experience here all that much richer. In spite of my nature as a worrier, she kept me on even ground, even when we were trapped in Europe due to a volcanic eruption. Though I leave her group for future endeavors, I'll always remember what she taught me; the most important of which was how not to say yes to every request that came my way. Thank you for everything you have done for me in my education, my research, and my life; I know the coming generations will be sure to echo my sentiment.

I also want to thank Matthew C. Fickus for his unofficial role as my co-advisor. His perspective as a mathematician opened new directions of research which have matured into the mathematical framework of this thesis. Between the disproportionately large burritos, tasty take-out, simulated rockstar fame, the hottest thing I've ever eaten, and the generally irreverent commentary that we both enjoy, it's easy to see why I did not mind making the trip to Dayton to work with him. The whiteboards in his office and in his home proved to be places of discovery for my work. In the same way he helped me develop a core component of this thesis, I know he will continue this tradition with those to come.

I would like to thank Markus Püschel and Gustavo K. Rohde for their vital contributions

as committee members and teachers. Markus' unique perspective on the intersection of mathematics and engineering inspired me in many ways; in particular, his lessons on how to write fast code have become one of the mantras I always preach. Gustavo's critical eye always revealed both the strengths and weaknesses of whatever he cast it upon, a key reason why this thesis stands where it does now. Their invaluable advice led me to improve the quality of my research by exploring new areas and considering new perspectives. The breadth and detail of this work was largely in part to their much appreciated comments and feedback.

I would like to thank John A. Ozolek and Carlos A. Castro, the two always impressively skilled pathologists whose original vision gave birth to this work. Their contributions and guidance helped not only this work, but myself grow into what this thesis represents. Along with providing me an incredible topic to work on, they gave me a new appreciation for the potential that collaborations between specialists can represent. I only hope that this work and what will come from it will help further their own goals just as their contributions furthered mine.

I also want to thank Melody L. Massar for helping me develop the mathematical framework and providing a system of checks and balances against the rather unique personalities of the three of us. You are fantastic mathematician, and assuming Matt does not send you into a burrito coma, I look forward to your success.

Thank you to group members, both past and present, for being fellow travelers down the same road. In particular, I want to thank Amina Chebira and Gowri Srinivasa for welcoming me to the group as if I'd always been part of it. I especially want to thank Amina for laying the foundation for this work which I hope will do justice to the quality of your own. For those that have only known me as the somewhat unorthodox senior student, you are in an incredible place and time in your lives, enjoy it to the utmost. Between CMU and Jelena, you will never find yourself wanting for inspiration, advice, and generally something to keep you occupied. May your journey be as fulfilling as my own.

I want to thank John W. Kelly and Chenlei Guo whose project in Jelena's course led to the first practical version of the histopathology vocabulary. In particular, I want to thank them for indulging my seemingly random idea that has now become a cornerstone of this thesis.

I want to thank Prof. Marios Savvides for opening the door to my Ph.D and guiding me through my first few years. His passion for research is something we should all strive for and I'm happy to know I've kept some of that with me. I cannot express my gratitude for your understanding when it came time to part ways as formal advisor and advisee. Thank you for your guidance and friendship throughout the years.

To the ECE department which has served as a second home for so much of my life, thank you for everything. To all of the faculty who have taught me, all of the staff who have helped me (mostly through indulging my constant appearances and pilfering of their snacks), and to my other acquaintances in the department, thank you for a wonderful 9 years as a formal student and a lifetime as a part of the ECE family.

To all my friends and other colleagues, both at CMU and otherwise, thanks for all the help and distractions that have kept me sane through these 5 years. As exciting as research can be, monotony will set in and it was thanks to all of you that I never found myself stuck for too long. To those of you with not so flattering, if not incriminating stories, let's save those for the reunions. Thanks for all the good times and best of luck to you all.

To my brothers, while we did not always see eye to eye on everything, my responsibility of being your older brother is something I am the better for. Though my choices of words and actions were not always ideal, I did strive to give you the benefit of my experience; I hope you found it useful in some way. I know that you will achieve your own brands of success to exceed my own. After all, we're brothers.

To my father, it has been a constant jest that I have followed in your footsteps more than my younger self might have ever have admitted or aspired to. Though we might laugh at this,

at this milestone, I can think of no one who has blazed a trail so bright and fill of promise for me to follow. As a researcher and a technical professional, I can express nothing but the utmost respect and admiration for all that you have accomplished. Your achievements have provided me an unfaltering beacon to follow in the best and worst of times. As you son, I could not have asked for a better father and what I am today is purely a reflection of the values you instilled in me. Thank you for all your support, love, and the wonderful opportunities that have allowed me to become the person I am today.

To my mother, you have always kept me grounded in what was truly important in my life, family. Knowing that you will always and unconditionally support me provides the foundation I need to achieve the excellence you alway wanted for me. I know that everything I have achieved and everything that is to come is only possible because of you. Just as constant as your reminders that there is always a seat at the dining table, a bed to sleep in, and a hug whenever I want, you will always be a source of support and love in my life.

Contents

List of Figures

List of Tables

List of Abbreviations

CAD   Computer aided diagnostic

CAT   Computed axial tomography

DFT   Discrete Fourier transform

DLM   Dead Leaves Model

DWT   Discrete wavelet transform

ES   Embryonic stem

FT   Fourier transform

GI   Gastrointestinal

H&E   Hematoxylin and eosin

HV   Histopathology vocabulary

LBP   Local binary patterns

LDA   Linear discriminant analysis

MPBM   Multiple-phase Boolean model

MR   Multiresolution

MRB   Multiresolution basis

MRF   Multiresolution frame

MRI   Magnetic resonance imaging

NN   Neural network

PCA   Principal component analysis

RGB   Red-green-blue

ROI   Region-of-interest

SPBM   Single-phase Boolean model

STFT   Short-time Fourier transform

SVD   Singular value decomposition

SVDD   Support vector data descriptor

SVM   Support vector machine

SVRDM   Support vector rejection decision machine

WHO   World Health Organization

Chapter 1

Introduction

The last time you visited your physician for relief from some ailment, he began by asking questions about your symptoms and likely causes of your illness. He probably also collected quantitative data such as your pulse rate and blood pressure. Based on the information collected, he prescribed some treatment course that was expected to cure your malady. However, the illness may have persisted in spite of the treatment. You returned to your physician for additional collection of information and an updated course of treatment. Until you were cured, you and your physician repeated these steps which represent the major divisions of medicine, diagnostic and therapeutic medicine (see Figure 1.1).

Diagnostic medicine aims to evaluate the illness that afflicts a given patient. This most often includes identifying the type of illness, ascertaining its severity, and determining its cause by utilizing a variety of techniques and tools. Such tools include standard patient questioning and collection of vital statistics but also include many more advanced methods such as X-rays, electroencephalography, and magnetic resonance imaging (MRI) to name a few. Development of new techniques and tools and improvement of existing ones is a major focus of both research and industry.

Based on the information collected by diagnostic methods, therapeutic medicine's goal is to prescribe the most effective course of treatment for the illness. Treatments are generally
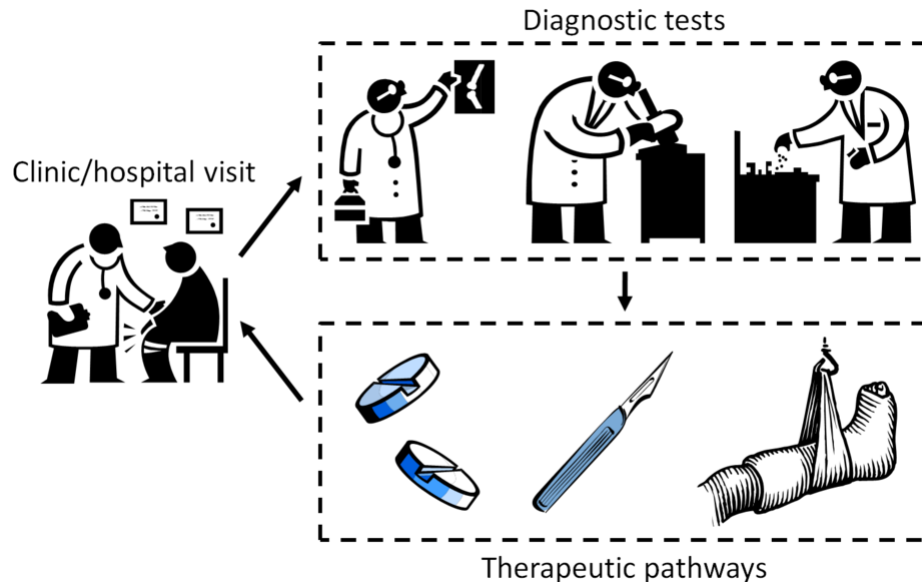
Figure 1.1: General medical pathway demonstrating the closed loop nature of diagnosis and therapy.

deemed effective based on a balance between their overall success rate, cost, and risk. Drugs are typically considered a much cheaper and less risky option when compared to surgery. Naturally, prescribing more expensive and riskier treatments occurs only when all other possible options have been exhausted. However, this decision is always based on the information provided by the diagnosis as more severe diagnoses likely demand more immediate and significant action.

While we have thus far presented information in the context of clinical medicine, medical research has as strong a connection to diagnostic and therapeutic medicine as clinical medicine. Medical research is of course the source of many of the tools and techniques used in the clinical domain but many other topics are also addressed. Such topics can be generally considered basic science questions. For example, understanding and controlling normal and abnormal development of humans is dependent on not only assessing development but also on applying a variety of procedures and observing their impact on development.

These two branches of medicine are developed and used hand in hand in every medical scenario. Thus, a meaningful contribution to either of them has the potential to be

far-reaching. It is in the diagnostic branch where we propose to make a contribution by developing a tool in the area of histopathology. To aid physicians in making accurate and reproducible diagnoses, we propose to:

Create a framework and methodology for the automated identification and delineation of tissues and their pathologies in hematoxylin and eosin stained images.

1.1 Pathology

Traditionally, the four major components of diagnostic medicine are anatomy, physiology, pathology, and psychology. Anatomy and physiology refer to the structure and function of the human body respectively. Pathology focus on understanding what, how, and why anatomy or physiology may be affected. In the context of medicine, pathology asks how illnesses manifest themselves by negatively impacting the form and function of the human body. To complete the description of diagnostic medicine, psychology addresses the thought and behavior of a patient. We will focus solely on pathology and more specifically histopathology.

At a coarse level, pathology focuses on understanding the causes and development of diseases while observing how cells are affected and in turn how these changes manifest as clinical symptoms. Generally, there are thought to be two main types of pathology, anatomical and clinical. Anatomical pathology observes changes in the macroscopic, microscopic, chemical, and molecular presentation of organs, tissues, and even the entire body. Clinical pathology utilizes laboratory analysis of bodily fluids such as blood through chemistry, microbiology, and other means.

Those that pursue education and training in this field are termed pathologists and are among the most capable diagnosticians in the modern medical system. Pathologists at large are expected to have a detailed knowledge of general human anatomy, physiology,

and pathology but many specializations exist. These specializations typically correspond with specific anatomical or physiological systems. For example, dermapathology focuses on the diagnosis of diseases of the skin while hemapathology addresses blood-related diseases. Extensive education and years of experience result in specialists whose diagnostic abilities are highly valued and often absolutely necessary in the overall course of treatment for many illnesses.

1.1.1   Histopathology

A specific and extremely useful subfield of anatomical pathology is that of histopathology. Histopathology by definition is the study of disease through microscopic examination of tissues and their component parts. It is a vital component in the overall clinical pathway of many diseases and is an important tool in the research of many medical topics. The overall procedure involves acquisition of a tissue sample or biopsy, chemical fixation, histological sectioning and mounting, staining, and visualization/imaging.

Standard Procedure

Acquisition is usually accomplished through surgical means that vary according to the anatomical region and specific need. For example, in the case of brain, breast, or skin tissue, a potential acquisition method is to insert a fine needle into the tissue and extract a sample through aspiration (suction) as shown in Figure 1.2(a). Another example is the endoscopic biopsy that we will discuss in detail in Section 2.1.2. As one may expect, the wide variety of acquisition methods will yield many fundamentally different types of samples. Additionally, the operating surgeon's abilities and patient's condition will cause samples acquired with the same method to present variations ranging from subtle to extreme.

Once the biopsy is acquired, it undergoes the process of chemical fixation. A simple description of this process is the replacement of all water in the biopsy with a chemical

(a) Biopsy

(b) Chemical fixation

(c) Histological sectioning and mounting

(d) Staining

(e) Visualization/Imaging
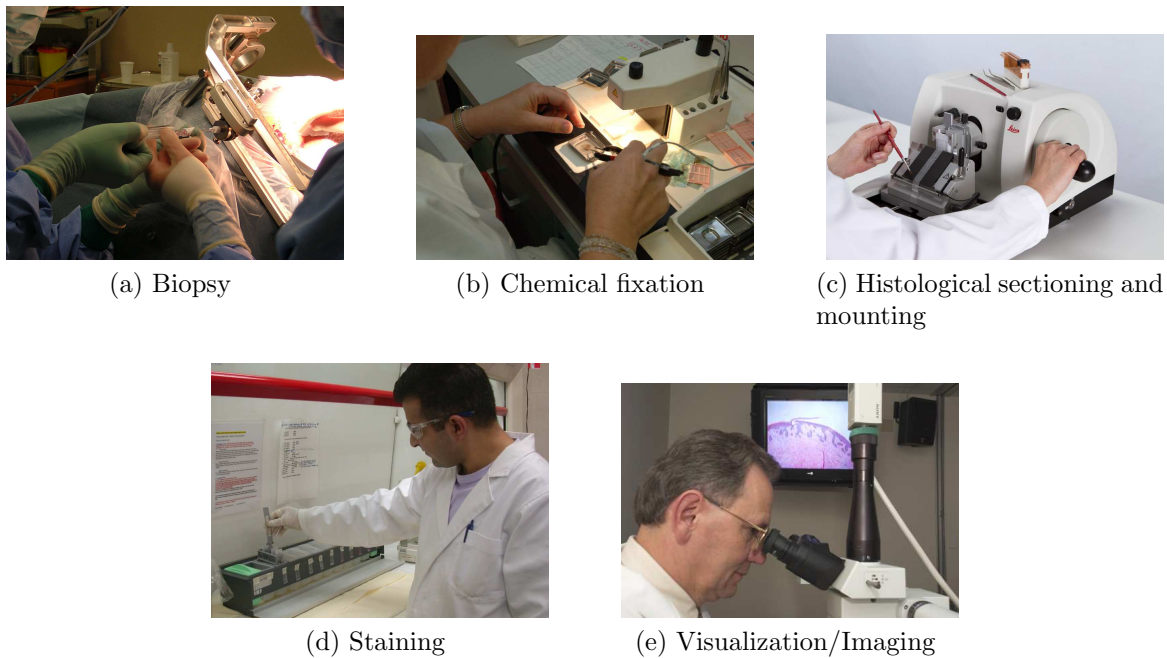
Figure 1.2: Illustrative examples of steps of histopathology standard procedure: (a) biopsy (www.wikimedia.com), (b) chemical fixation (www.protocolsonline.com), (c) histological sectioning and mounting (www.leica-microsystems.com), (d) staining (www.bio.mq.edu.au), and (e) visualization/imaging (www.pathology.washington.edu).

reagent of choice, typically paraffin. Effectively, the paraffin fixation will turn the previously soft biopsy hard allowing for the next steps in the process. During the acquisition stage, the tissue will have suffered some damage and may have even effectively died but individual cells may still be functional. However, after the chemical fixation step, the tissue is in fact dead and suspended to prevent decay. As a result, a fixed biopsy can be stored indefinitely for future use but no time-dependent behavior of the tissue can be observed. The final part of this step is to take the now fixed biopsy and embed it in a block of paraffin as illustrated in Figure 1.2(b).

At this stage, the fixed biopsy is a three-dimensional object that (in most relevant cases) will allow for some direct microscopic examination as is, specifically that of its exterior. However, the interior of the biopsy, which is often very important, cannot be seen at this stage. Furthermore, the ability of high-quality microscope to provide accurate visualizations

is dependent on its ability to optically focus on a very narrow depth of field. Additional issues related to transmission of light through non-air mediums further confound the direct microscopic examination of the fixed biopsy. As a result, very thin slices of the biopsy must be made to yield precise visualization. A tool called a microtome, an example of which is shown in Figure 1.2(c), is used to slice the paraffin block into serial (consecutive along the depth of biopsy) slices whose thickness is on the order of a few micrometers which are then mounted on slides. Naturally, the quality of the microtome and the operating technician's ability will affect the quality and consistency of the slices.

The slides created at this point could be observed under a microscope but the features of interest are not likely to be well defined. Specifically, as a result of the chemical fixation, the mounted slices lack distinct coloration. While some structure might be observable, much more information can be obtained by applying specific chemicals called immunohistochemical stains to the slides. Each stain is designed to bind to certain characteristic proteins in tissues. Since the presence of particular proteins correlates very strongly with specific tissue components, the tissue structure is made visible by this process. The choice of stain dictates which structures are highlighted. Furthermore, combinations of stains can be used to effectively color a variety of structures and help differentiate between them. Though the exact procedure for each individual and combination of stains varies, we present an illustrative example in Figure 1.2(d). The gold standard or most routinely used stains are those of H&E that we will discuss in more detail in the next section. As there are many manufacturer's of stains, many staining protocols available, and the variation introduced by the technician, the exact staining amongst a set of samples is likely to vary in many situations.

Once all the previous steps have been completed, that pathologist is now able to observe the slides under a microscope as shown in Figure 1.2(e). His assessment of the observed tissues will be guided not only by his knowledge and experience, but also by the situation at hand. For example, knowing that he is to evaluate the sample for a particular disease allows

a pathologist to focus his efforts for a much more precise and efficient evaluation. Direct observation through a microscope is almost always done initially but thanks to the advent of digital sensors and their integration into microscopes, images can also be taken for future use. As would be expected, additional source of variation such as a pathologist's preferences, microscope optical quality, and the quality of the image sensor influence the resulting image.

Hematoxylin and Eosin

As mentioned before, the choice of stains used dictates the overall presentation of tissues in a sample. The most popular choice of stains in pathology is the combination of hematoxylin and eosin (H&E). Hematoxylin binds primarily to those structures containing nucleic acid such nuclei and ribosomes, coloring them in shades of blue and purple. In contrast, eosin binds to structures primarily made up of protein such as cytoplasm and red blood cells, coloring them in shades of red and pink. These complementary chemicals stain the majority of the structures present in tissue and result in a very informative coloration of a sample. We present some example images of samples stained with H&E in Figure 1.3.

These stains are permanent and as a result, various institutions around the world are in possession of immense collections of H&E stained samples representing many sources and large spans of time. The amount of samples available supports the claim that H&E staining is of vital importance to the medical system. It would not be unrealistic to claim that the overwhelming majority of pathology labs throughout the world are capable of performing H&E staining.

1.1.2 Role in Clinical Medicine

Histopathology, and more specifically, H&E staining is of vital importance to many medical pathways as a diagnostic tool. For example, in the United States alone, more than 15 million upper and lower endoscopies are performed each year. During many of these procedures,

(a)   (b)

(c)   (d)
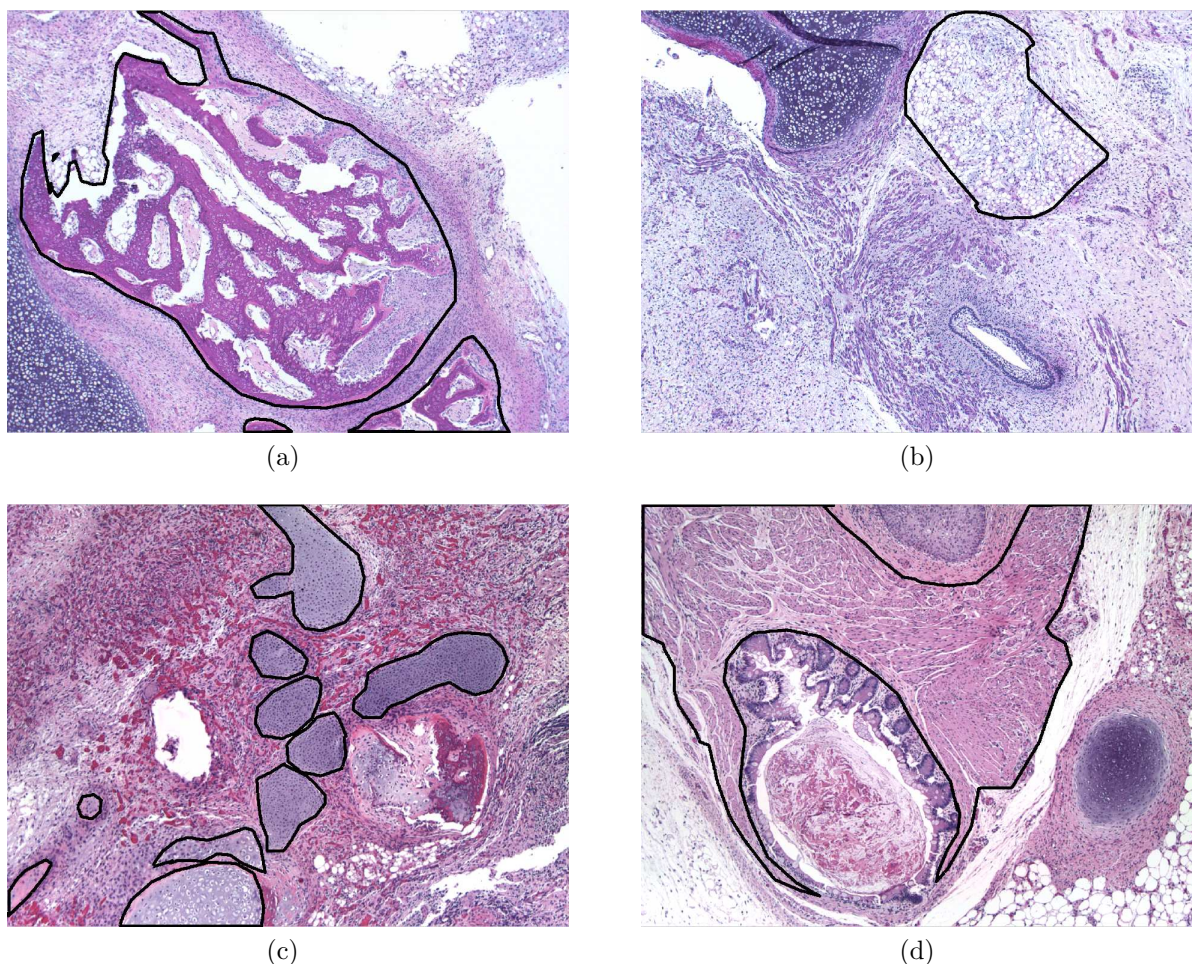
Figure 1.3: Example H&E-stained images with particular tissues outlined: (a) bone, (b) fat, (c) cartilage, and (d) smooth muscle.

biopsies of the GI tract are acquired and subsequently analyzed by pathologists. This particular screening procedure is vital in the detection of many illnesses such as colon cancer and inflammations.

Similar procedures exist for many different organ systems and their associated diseases but the general category of cancer is the most well-known. The term biopsy is most often linked to the diagnosis of cancer. Specifically, pathologists are able to identify the type and level of malignancy present in a given biopsy. These diagnoses become the cornerstone upon which all subsequent action is based. Slight changes in a diagnosis from one pathologist to another or from sample to sample can have far-reaching consequences. For example, given

some sort of grading scheme for a specific type of cancer, the assignment of a low grade
might lead to a treatment of just drug therapy. However, a moderate grade might require
more intensive action such as chemotherapy while higher grades may lead to surgery. The
oncologist, a physician who specializes in diagnosis and treatment of cancer, will decide
to prescribe any of these treatments based in part on the pathologist's assessment. As a
result, accurate diagnosis is critical to prescribing the appropriate course of treatment while
avoiding excessive and potentially harmful alternatives.

Cancer is a relatively extreme family of afflictions that are often diagnosed with histopathol-
ogy, but many less severe and more frequent maladies also benefit from this diagnostic
tool. Infections and inflammations of various organ systems can be identified and evaluated
through histopathology samples. Even some parasitic organisms can be clearly recognized
under H&E staining. Effectively, histopathology offers some information concerning any
disease that manifests itself in an anatomical or physiological manner at a microscopic level.

1.1.3   Role in Medical Research

Medical research in general utilizes most of the same diagnostic tools as clinical medicine
but not with the goal of prescribing treatment for patients. Rather, these tools are used to
provide qualitative and quantitative evaluations of data derived from experiments. For ex-
ample, while the biochemical analysis of blood may reveal much about a patient's condition,
the same analysis can be used to observe the effects of new drugs being developed. A gen-
eralization (that is not always true) of this concept is that clinical medicine uses diagnostic
tools to asses the condition of a subject prior to treatment while medical research uses them
to observe a subject's condition after treatment.

Histopathology is one such tool that is capable of providing many insights into a variety
of research applications. One such family of applications that we will discuss in more detail
in Section 2.1.1 is that of stem-cell research. Stem-cells are capable of becoming any cell in

the human body given the correct environmental conditions and stimuli. They hold great potential in clinical areas such as tissue regeneration and treatment of cancer. They also provide a means through which researchers can further our understanding of basic human development and more specifically the reasons for abnormal development. Much of this research has direct clinical impact but the potential to answer many basic science questions is equally significant.

1.1.4 Fundamental Tasks

The importance of histopathology is clear when one looks at its role in both clinical medicine and medical research. These applications may require very simple contributions from pathologists or may need very complex action on their part. Among the many different tasks that may be required by the multitude of possible applications, there are four fundamental tasks that the most critical to the vast majority of applications: tissue identification, pathology identification, evaluation, and delineation.

Tissue Identification

As pathology is based on the recognition of changes in tissues as a result of some condition, the first task must be to identify the tissue so as to provide a context for these changes. In some situations, the tissue is already identified as a result of prior knowledge, i.e., knowing that it was acquired from a particular organ or the like. However, in some situations (such as some research ones), the tissue identity is not provided. In such situations, the pathologist begins with the identification of the tissue or tissues present in a given sample.

Pathology Identification

Along with evaluation, pathology identification is the most common and most important fundamental task for a pathologist. While pathology refers to the general diagnostic field

we are discussing, it also refers to the condition of a given sample and in particular how it deviates from the normal or healthy condition. Identifying the pathology present in a sample has obvious importance as the first step in diagnosing a disease in a clinical setting or understanding the results of some experiment in a research environment. At its coarsest level, pathology identification is the task of distinguishing between normal/healthy tissue and abnormal/diseased tissue.

Evaluation

Once the pathology of a tissue has been identified, it is common for that pathology to be evaluated in terms of progression or other important characteristics. The clinical diagnosis of a disease will include an evaluation of how advanced the onset of the disease is. As stated before, this evaluation has direct consequences on the exact treatment that will be prescribed and the risks associated with it. In a research application attempting to cure some ailment, determining the efficacy of the drug is done by evaluating how reversed the progression of the ailment after application of the potential cure. Naturally, the level of success with any form of the cure will drive or discourage further development of the current design. This task is also the most subjective as a pathologist's training, experience, and preferences directly influence the evaluations they provide possibly resulting in both subtle and gross differences with other pathologists' evaluations.

Delineation

This task is a possible combination of the previous three tasks and is mostly relevant in research applications. Delineation is the process of not only determining which tissues and/or pathologies are present in a sample/image but also where they are spatially. This is effectively a segmentation of the image into regions with clear identities. In many clinical applications, it is not so common to consider more than one tissue or pathology and thus delineation is

not always required. However, in many research applications, the identities and locations of tissue and pathologies can be vital to answering the research question of interest.

For example, consider tissue regeneration research where measuring how much the tissue of interest has grown over time is a critical indicator of progress. In this scenario, the spatial location and size of the tissue of interest must be measured. It is this task that we are proposing to automate for a variety of applications through one unified algorithm. Furthermore, delineation tasks are complicated by the uncertainty in the exact position of boundaries between regions (tissues/pathologies) of interest. This can be an especially troublesome point in the design of algorithms attempting to automate this task in some capacity. Though this is a fundamental task, it is not a simple task due to the complex challenges associated with it.

1.2   Digital Pathology

The impact of histopathology, and pathology in general, is immense in both clinical medicine and medical research. As a result, there is a constant and continuing effort to improve this general area not only by increasing its efficacy through science and technique, but also by incorporating and developing new and relevant technology. Perhaps the first such modern technology was the advent of the digital sensor with which pathologists are now able to take high-resolution images of samples for distribution and storage. While the digital image provided a more universal and more easily transferred medium for data, other new technologies focusing on improved throughput and integration have been developed.

The advent of these technologies has begun with what many refer to as the digital pathology revolution; the transformation of the standard pathology model into one that seamlessly integrates modern technology, specifically the modern computing infrastructure and high-quality automated instruments. While many individual efforts are contributing to this revolution in different ways, they can generally be categorized as contributing to: improved

digital imaging high-throughput scanning, integration with modern networking, and image analysis software.

Improved Digital Imaging

Efforts in this area primarily focus on increasing the effective resolution of the acquired images with respect to the physical dimensions of the samples or in other words, increasing the number of pixels per unit area. Increasing the resolution of the digital sensor is a simple solution and is appropriate as the corresponding industry develops more cost-effective products. Improvements in the optical elements of the imaging system, i.e., the lens of the microscope, are also pursued but mostly independent of the specific needs of pathologists.

Another concern is the overall field of view acquired by these sensors. Depending on the required amount of detail, images are often acquired at fairly high magnifications resulting in images which represent only small portions of the entire sample. In certain applications, large areas must be imaged at such resolutions for proper evaluations to be made. As one might take a series of images to represent a panorama, the same must be done in a much more rigorous fashion in pathology applications. While this could be performed manually by the pathologist, it would be subject to human error. To address these issues, companies such as Leica have developed robotic imaging systems capable of precisely imaging these large areas at high resolution. These products are referred to as whole slide imaging platforms with an example being shown in Figure 1.4(a).

High-Throughput Scanning

The number of pathology labs around the world results in a vast amount of samples being generated including those mounted on slides. In the typical workflow, these slides are viewed and imaged under a microscope. Furthermore, this was previously done in a completely manual fashion by pathologists and required a significant amount of time and effort. Thanks

(a)

(b)
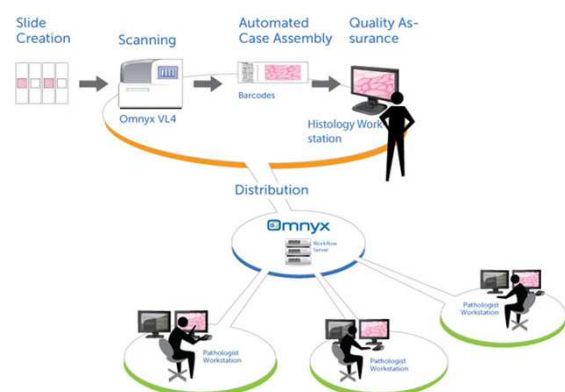
(c)

(d)

Figure 1.4: Examples of four main digital pathology areas from industry: (a) whole slide imaging (www.leica.com), (b) automated scanner (www.aperio.com), (c) integration with modern networking (www.omnyx.com), and (d) image analysis software (www.omnyx.com).

to advances in many fields, hardware is available to automate this process such as the example product shown in Figure 1.4(b). Not only do such products improve the speed at which a single slide may be imaged but they also increase the number of slides that may be imaged at one time. While the term high-throughput scanning applies to many different non-pathology related fields, it is appropriate here as the overall throughput of the system in terms of slides being imaged is vastly improved in comparison to manual methods.

While the exact design and capabilities of these scanners vary, the vast majority of

modern products emphasize the ability to image large numbers of slides without human intervention. This means that technicians who would otherwise be occupied by imaging individual slides one after another can spend more time on providing high quality slides through more precise execution of other steps in the standard procedure. Integration of these machines into existing labs is becoming more and more common resulting in even more data being generated. This ever-increasing amount of data is a critical argument for the development of image analysis tools such as the one this thesis presents.

Integration with Modern Networking

As is the case in many fields, collaboration between experts is encouraged to advance the area. In medicine, these collaborations often come in the form of specialists offering their opinions on a given case in the hopes that their collective knowledge will yield a better result. In pathology, such collaboration requires the sharing of data amongst each other; a task for which modern networking can be of great use.

The main focus of integration of modern networking into the pathology workflow is making data easily accessible from any location regardless of where it may physically be stored as shown in Figure 1.4(c). Effectively this means storing pathology data on a networked device and providing remote access to designated users. This could be limited to the internal network of a single hospital or could be as vast as a global repository of information accessible by all pathologists. By making data available, many pathologists can contribute their expertise to a problem.

Image Analysis Software

The digital image is fast-becoming the fundamental unit of data in pathology and in turn, software capable of providing advanced analysis of these images is being developed. Specifically, researchers are creating automated (independent of human action) algorithms capable

of performing tasks normally done by pathologists manually. Many computer aided diag-
nostic (CAD) tools aim to detect the presence of particular diseases and assigning levels of
severity. Other algorithms may be more focused on enhancing or modifying the image in a
way that makes others tasks easier. Some software aims to extract information that is not
intuitive or easy for humans such as precise quantitative measurements.

Fundamentally, image analysis software aims to reduce the work load of pathologists
by introducing some level of automation for certain tasks and performing other tasks that
are beyond the capabilities of humans (Figure 1.4(d)). This thesis proposes such a tool to
partially automate a wide variety of simple but very important tasks and in doing so, provide
a platform for other algorithms. In the next section we will expand on this general area and
further motivate its development.

1.3   Benefits of Automated Image Analysis

At its essence, automated image analysis addresses some basic limitations of expert (manual)
analysis. While analysis by trained and experienced pathologists is incredibly accurate and
will be the benchmark for performance for some time to come, it does come at a cost.
The ever increasing demand for high-quality and fast analysis has created a burden on
pathologists that is not sustainable. Furthermore, the increasing amount of collaboration
between pathologists in a variety of tasks has created new challenges that must be addressed.
After careful consideration, we believe the primary limitations to expert analysis to be that
it is time-consuming, difficult to reproduce, inconsistent between experts, and difficult to
extract some quantitative information.

Time

The first limitation is rather intuitive given the ever-increasing amount of data that pathol-
ogists must analysis. For example, it is estimated that as part of colorectal cancer screening

about 14.2 million endoscopies were performed in the U.S. during 2002 [1]. In reality, not all of these procedures involve biopsy acquisition and analysis but in order to maximize the likelihood of detection and evaluation of serious illnesses, they would. In this situation, based on the assessments of Drs. Castro and Ozolek, about 10 minutes are needed on average to analyze each such biopsy in sufficient detail. Based on the one estimate of the median annual income of all U.S. pathologists [2] of $220,000 for about 2,400 hours per year, the cost of analyzing all such endoscopic biopsies from just pathologist compensation is approximately $216.9 million/year. Clearly, this simple analysis shows that the financial burden can be significant and becomes even more meaningful when one starts to consider all the other such applications, some of which with even larger prevalence and many with greater associated costs.

Automated image analysis can help to reduce the overall commitment of time in many ways. In the context of the colorectal cancer screening discussed above, consider the following hypothetical example. A typical pathology department of a hospital will be staffed by a combination of a few very experienced pathologists and many relatively inexperienced ones. Under the standard operating paradigm, endoscopic biopsy samples from colorectal screenings are evenly distributed amongst all of these pathologists regardless of experience. As a result of this workflow, some undesirable situations may arise. First, experienced pathologists will have to spend time identifying and confirming cases where there is clearly no presence of colorectal cancer, a relatively simple diagnosis to make. Second, the less experienced pathologists will likely have to address very difficult cases where the exact degree of colorectal cancer is difficult to assign and may make a mistake in doing so. Lastly, in order to analyze all assigned samples, the pathologists will be forced to commit an average amount of time to each sample regardless of difficulty which will likely be less than optimal in terms of accuracy and efficiency.

If an automated tool capable of assigning perceived levels or confidences of normalcy

(absence of colorectal cancer) to any given sample was available, the distribution of samples could be done in a much more efficient manner. One option is that every sample would be evaluated by this algorithm and samples for which there is overwhelming confidence of normalcy would be given to the less experienced pathologists. At the same time, those samples not yielding high confidences of normalcy would be sent to the very experienced pathologists. As a result of this sort of allocation of samples, the primary role of the less experienced members of this department would be to confirm the perceived normalcy of the samples given to them and to correct mistakes made by the automated tool. The more experienced members would likely have to address a decreased number of cases and could thus spend more time delivering more accurate diagnoses. The overall result would be reduction in the overall time required for analysis and an improvement in the accuracy of diagnosis, especially for more complicated cases.

Reproducibility

This limitation is merely a consequence of the natural variation we as humans demonstrate in all activities. Regardless of how often we may repeat a given task, minute variations in our exact performance will occur. The total effect of these variations range from the entirely inconsequential to those completely dictating the outcome of the activity. The same considerations are true when a pathologist analyzes a given sample whether it be in a clinical setting or during some sort of research. For example, in a clinical setting, the exact evaluation of an identified ailment dictates the course of treatment. Though it will not be common, if a pathologist assigns different evaluations to multiple samples (whether they be from the same or different patients) that are in fact of the same severity, the implications may be great. Along with accuracy, consistency is a desired quality of any diagnosis.

The introduction of automated image analysis to these scenarios can clearly help mitigate such issues. While some variation might be expected in the analysis provided by such

tools as a result of their particular design (i.e., some sort of intentional randomness), most algorithms are expected to deliver reproducible results. When presented the same piece of input data repeatedly, the same general result is expected each time. The positioning of such algorithms to help introduce consistency is entirely dependent on the particular workflow being addressed. One possibility is to place such an algorithm prior to expert analysis so as to provide a consistent starting point. Alternatively, the algorithm may be positioned after the expert's role in order to help correct for the variations that may occur. Ideally, such tools will function as purely objective evaluators of data that can be relied upon to be consistent.

Consistency

Differences in opinions between experts is common in any field, the analyses of pathologists being no exception. As pathology grows in importance and various forms of communications technology make it easier for pathologists to collaborate, the analysis of data will now depend on the collective efforts of many experts as opposed to the individual few. In the clinical domain, the overall accuracy of a diagnosis will be improved by leveraging a greater collective sum of knowledge and experience. Similarly, in research, more confident conclusions can be drawn by eliminating biases that may exist in individuals. However, even with such desirable effects, these majority opinions will sometimes suffer from strong disagreements that may need to be mediated in some fashion. A stereotypical example is that of the two highly experienced and well respected pathologists whose individual analyses may come into conflict due to their personal preferences.

Similar to the previous limitation, automated tools act as objective and reproducible evaluators of data. When considering the disagreements that may arise between experts, the automated algorithm can act as a relatively impartial mediator. We say relatively as the manner in which the algorithm was designed and put into use has an obvious effect on

the general nature of its results. Still, it may help mitigate strong disagreements between experts by providing them a common frame of reference for their results and perhaps lead to an agreement amongst them. Such resolutions may be as simple as the experts agreeing that the algorithm's evaluation which exists somewhere between their opinions is more correct than either individual one. On the other hand, it may lend more credence to one opinion than the other and thus help convince one party to change their evaluation. Regardless, such algorithms can in fact act as negotiators in these situations where experience and personal bias have a dramatic impact on any collaboration.

Extraction of Quantitative Information

This limitation is purely a function of a human's capacity to precisely quantify various aspects of images. For example, nobody can be expected to precisely describe the average color in particular regions of an image let alone the image as a whole. A human may be able to provide general qualitative assessments such types of colors but not precise measures. Similarly, when asked for the overall size of a region or how much it neighbors another particular region, precise and accurate measurements are difficult to give. A more extreme example is the counting of a particular structure and a quantification of their organization in some manner. Naturally, this is exactly the type of work that computational algorithms are well suited for.

The need for such information is greater or lesser depending on the application. For example, in clinical scenarios where mere detection of an illness is the goal, precise quantitative information may not be required. In the same setting, evaluation of the severity of the illness may be dependent on a variety of quantitative measures. The vast majority of research applications will require some form of quantitative assessment to provide evidence for whatever conclusions they may draw. Equally important as easing the workload and improving overall effectiveness, the ability to extract quantitative information from data is

a primary motivation for utilizing automated image analysis algorithms.

1.4 Pathology Image Analysis Related Work

The onset of digital pathology is directly supported by the advancement of technology developed in the four main areas mentioned previously. Of these four areas, the development of image analysis tools has been greatly supported both by industry and academia. As a result, there is already a large amount of existing work done resulting in many novel and useful image analysis tools. While there are many questions and problems still to be addressed, some specific sub-areas have received much attention by the research community. Specifically, we refer to cancer related applications, image retrieval systems, and data normalization algorithms.

We begin this discussion by commenting on the application specific nature of much existing work and what it implies for the general area. We will also discuss some existing works representative of the field at large. Our focus will be on their intended applications and some of the primary contributions and conclusions they present. While the discussed works do represent a broad cross-section of published research they do not represent all such work.

Application Specificity

It is natural to expect that when trying to contribute to the solution of a problem, your contribution is specific to that problem. In the context of developing image analysis tools, these problems most often are constrained to specific applications. For example, the development of a tool to assign a Gleason grade to a prostate cancer sample. The application in this example is very specific in that it addresses a specific organ and has a specific task unique to this organ. This type of specificity in goal is very useful in developing effective tools, but it comes at a price.

Given the immense number of possible applications, all of which are related through

their use of histopathology, is it wise to develop individual solutions to each problem even though there may be strong connections between them? It would be naive to claim that when presented with a new task, all previous knowledge is discarded including solutions to related tasks, but it is not often that two different applications attempt to leverage the same solution. As we present some related work, the application-specific algorithms will not only present their usefulness but also their limitations with respect to cross-application applicability. Thus, it is our opinion that a more lasting impact could be achieved by trying to understand the similarities between applications and how a single given algorithm can leverage them to achieve many different tasks.

Perhaps the most convincing argument for this opinion is the capability of the trained and experienced pathologist. While not every pathologist is capable of every histopathology related task, when presented with most image analysis related problems, their ability to accomplish such tasks is extraordinary. This seems to suggest that while there is clearly some specialized knowledge relatively unique to any given application, there is a large body of knowledge common to many applications. In the same way that pathologists do not discard their existing knowledge and experience every time they are presented with a new task, we believe an algorithm can be designed to have similar capabilities when addressing a large body of histopathology related applications. This is a notion that we will discuss in more detail throughout this thesis.

Cancer Diagnosis

As one of the leading concerns of modern medicine, cancer of all sorts is a popular application when developing image analysis tools. In 2007, cancer collectively was the 2nd leading cause of death in the U.S. [3] with heart disease being the leading cause. The American Cancer Society estimated that in 2010 alone, there was to be an estimated 1.5 million new cases of cancer in the U.S. alone [4]. As a result of this high onset of cancer in the U.S. in 2010,

about 570 thousand cancer-related deaths were expected, nearly 1 in every 4 deaths [4]. These statistics emphasize the importance of this family of disease and also the burden which it places on the medical system and society at large.

The risks and costs associated with cancer treatment encourage aggressive research into providing more effective tools not only for therapy but also for diagnosis. Faster and cheaper diagnostics would allow for more people to be screened more often with the expected benefit of higher likelihood of early detection, considered a key factor in successful treatment of many cancers. The specialization of pathologists with respect to certain types of cancer mean that there is limited access to these professionals. Algorithms capable of mimicking some of the more basic tasks performed by such specialists would increase access to high-quality treatment. The limited number of experts also places an ever-increasing workload on them as various factors may lead to increasing prevalence of certain cancers. All of these reasons and more provide a strong argument for the continuing efforts in developing algorithms and tools for the diagnosis and treatment of cancer.

The higher incidence of certain types of cancer has led to greater efforts in addressing tasks related to those types as opposed to others. For example, breast and prostate cancer are the most common type of new cancer cases in women and men respectively in 2010 [4]. They are also the 2nd leading cause of cancer-related deaths with lung cancer as the leading source [4]. It is expected that 1 in 8 Caucasian women will develop breast cancer with 1 in 3 Caucasian men developing prostate cancer [5]. From these figures, one can appreciate the impact of these two specific types of cancer alone. Similar statistics help justify the increased or diminished attention and effort put forth in developing tools for particular types of cancer. These tools address the detection of cancer (normal/benign versus cancerous/malignant tissue), grading of cancer following established medical protocols, and image retrieval for case cross-referencing.

Detection of Cancer Algorithms for the detection of cancer are most often automated ones that aim to reduce the burden of work on pathologists though some take advantage of human intervention to improve and validate their results. It is an important distinction as there is a current expectation that no algorithm will ever be the final prescriber of treatment not only as a result of government regulations but also due to society's perception of how medicine should be administered. Regardless, as an actual tool or mere proof of concept, many works focus on algorithms for cancer detection [6–18]. It should be noted that cancer detection is merely the recognition of some level of malignancy in a given sample; this detection is often paired with a grading step depending on the application.

Grading of Cancer Assigning a level of malignancy or grade to a cancerous sample is nearly as important as detecting the cancer in the first place. Just as you only require a band-aid for a small cut as opposed to stitches for a large gash, the grade of cancer dictates how aggressive the required treatment will be. Cancer treatment is the result of a risk-benefit analysis as while various treatments offer improved chances of eliminating the cancer, they go hand in hand with higher risks of side-effects. The choice to pursue potentially more beneficial but riskier options is directly based on the severity of the cancer and the resulting need to act. Thus, accurate grading of cancer is vital to prescribing the appropriate level of treatment, one that will not cause more harm than good. Much work has and continues to be done to aid in this vital task [6–8, 10, 11, 15, 16, 18–20].

For this reason, cancer grading is often standardized by establishing grading protocols specific to various types of cancer. One of the more well-known examples is the Gleason grading scale associated with prostate cancer [21]. Such protocols provide a standard with which pathologists can universally convey their appraisal of the cancer severity. Naturally, person-to-person variation and inconsistencies in the protocol exist, but it is a much preferred alternative than the solely personal evaluation that would be done otherwise.

For example, the works of Doyle, Naik, and Madabushi [6–8] address the task of breast

and prostate cancer detection and grading. More specifically, when addressing prostate cancer, the standard Gleason grading scale is used. This is in contrast to a less formalized grading in breast cancer provided by their collaborators. In these works, the authors propose a series of textural and architectural features to quantify the difference between normal and cancerous tissue. Furthermore, they use various embedding techniques to automatically learn an order to these features that accurately reflects cancer severity. These works represent an incremental evolution of these basic ideas resulting in an effective algorithm for the task of prostate cancer detection and grading.

The work of Glostos et al. [15] proposes an algorithm for grading astrocytomas (a type of brain tumor) based on the World Health Organization's (WHO) grading system for tumors of the nervous system [22]. The WHO grading system is based on mostly qualitative criteria such as the morphology and distribution of key types of cells though some qualitative criteria concerning patient history are also present. In developing their algorithm, the authors guided some of their design on the criteria proposed by the grading system. Specifically, size and shape of nuclei along with DNA and chromatin distribution were encoded by using a variety of morphological and textural features. Automated grading was accomplished by learning the separation of low and high grade cases in the feature space via a support vector machine (SVM). As a result, they developed an encouraging algorithm that already fits into the known scheme of brain tumor treatment.

Image Retrieval As mentioned before, the prevalence of pathology labs around the world and the advent of digital pathology has led to a massive increase in the amount of data collected and analyzed. This abundance of information has many uses but one of the more prominent ones is the possibility to compare data against each other to provide a precedent for the appropriate treatment. Basically, given that a given condition has been treated in the past, either successfully or unsuccessfully, should another example of this condition arise, a physician can reference these past cases for guidance concerning treatment. However, unless

this vast collection of past cases is very carefully labeled and indexed, the task at hand is to determine which past case is the most similar to the current case. This is the task of image retrieval in the case of histopathology related applications.

Image retrieval is a specialized case of the more general task of image matching, a common task in research concerning search engines for example. Some key considerations in image retrieval is the possible presence of a priori information. As an example, a pathologist will most likely know which organ system and perhaps which condition is to be present in a given sample. Based on this knowledge, the amount of data to be searched is smaller than it would be otherwise and consequently should make the task easier. Furthermore, visual similarity is not necessarily sufficient in image retrieval tasks. While performing image matching, it may be enough to find two an image in the database that looks similar to the reference image. However, in a specialized image retrieval task, the level of visual similarity might not accurately represent the desired similarity between to samples. For example, a pathologist may consider two images that look completely different similar because they present similar structural and architectural cues as opposed to visual ones.

Such specialized systems are being designed for a variety of purposes [23, 24]. These systems often leverage the context of the application in order to simplify the problem. This can be in the form of some specialized meta-information available or application-specific image descriptors. For example, if the task is to retrieve cancer samples with similar levels of malignancy, the number of nuclei may be a good descriptor in this context. As a common theme, many solutions take advantage of the specifics of the application but in doings so can be limited in their overall applicability or usefulness.

Data Normalization The last major topic we will discuss concerning pathology related work is that of data normalization. As a result of the sheer number of pathology labs, techniques, technician skill and preference, and data sources, histopathology data will contain many variations. Some of these variations will be as obvious as physical artifacts and defects

while others are more subtle such as variations in staining coloration. Ideally, when analyzing multiple samples of the same tissue subject to the similar conditions, these multiple presentations will be similar in appearance. However, there can be large variations amongst these samples for many reasons.

To address these variations, some work has been done to normalize data with respect to possible sources of change. For example, Mecenko et al. [25]) present a method for stain normalization based on demixing methods. Specifically, it is assumed that prototypical stain colors exist in a given sample and that all colors in the sample are merely weighted linear combinations of these prototype colors. A singular value decomposition (SVD) is applied to the data, after some pre-processing, to determine these prototypes. Given multiple images, each with their own prototype colors, normalization amongst them can be achieved by aligning the colors with each other.

1.5   Our Goal

We have so far explained histopathology and the impact it has as result of its use in a variety of important applications. In this context, we have also discussed our opinion that while application specific work is useful, it does not necessarily advance the overall area of image analysis in the histopathology domain in a meaningful way. As a result of the clear need for such image analysis algorithms and our desire to make a significant impact on the area as a whole, we propose to:

Create a framework and methodology for the automated identification and delineation of tissues and their pathologies in H&E-stained images.

1.5.1   Motivation

The importance of histopathology to both the clinical and research domains of medicine provides more than sufficient justification to pursue related research. The development of tools, specifically image analysis software, can have a great and perhaps immediate impact on many important areas. By reducing the burden of time and effort on pathologists, the quality of care can improve through reduced costs, more accurate diagnoses, and faster throughput.

The motivation for this goal stems from the desire to make a more broad contribution to image analysis tools for histopathology than any one application specific algorithm. As discussed before, much existing work begins with a specific application in mind and proceeds to create algorithms within the context of that application. That specificity is useful in that it limits the scope of the task and can lead to very effective algorithms. However, the re-iteration of this process for every new application begs the question as to whether or not there is a more unified solution to more than one of these applications.

The capability of expert pathologists to perform so many tasks in so many different settings without re-training suggests that they have some level of universally applicable knowledge. Of course there are situations in which additional education may be required to achieve a truly expert level of competence but for the most part, it is our observation that many basic tasks are well within their grasp. Based on this observation we will develop an algorithm capable of mimicking this cross-application capability demonstrated by pathologists.

1.5.2   Challenges

The very nature of histopathological data presents many challenges in both the clinical and research domains. Before speaking about domain specific challenges, let us first consider the variations that arise from the acquisition procedure itself.

Beginning with the process of acquiring the biopsy, e.g., surgery, the exact environmental

and patient conditions can alter the sample greatly, For example, consider biopsy acquisition during an endoscopy. In this situation, the sample is acquired by cutting away a small piece of the inner lining of the GI tract. As the GI tract is essentially a tube, the angle of the cut made strongly influences the presentation of certain characteristic structures in the tissue. Figure 1.5 presents some examples of such GI tract biopsy acquisitions. While the surgeon will take care to be as consistent as possible in this regard, it is very difficult to achieve total consistency,

(a)

(b)

(c)

(d)
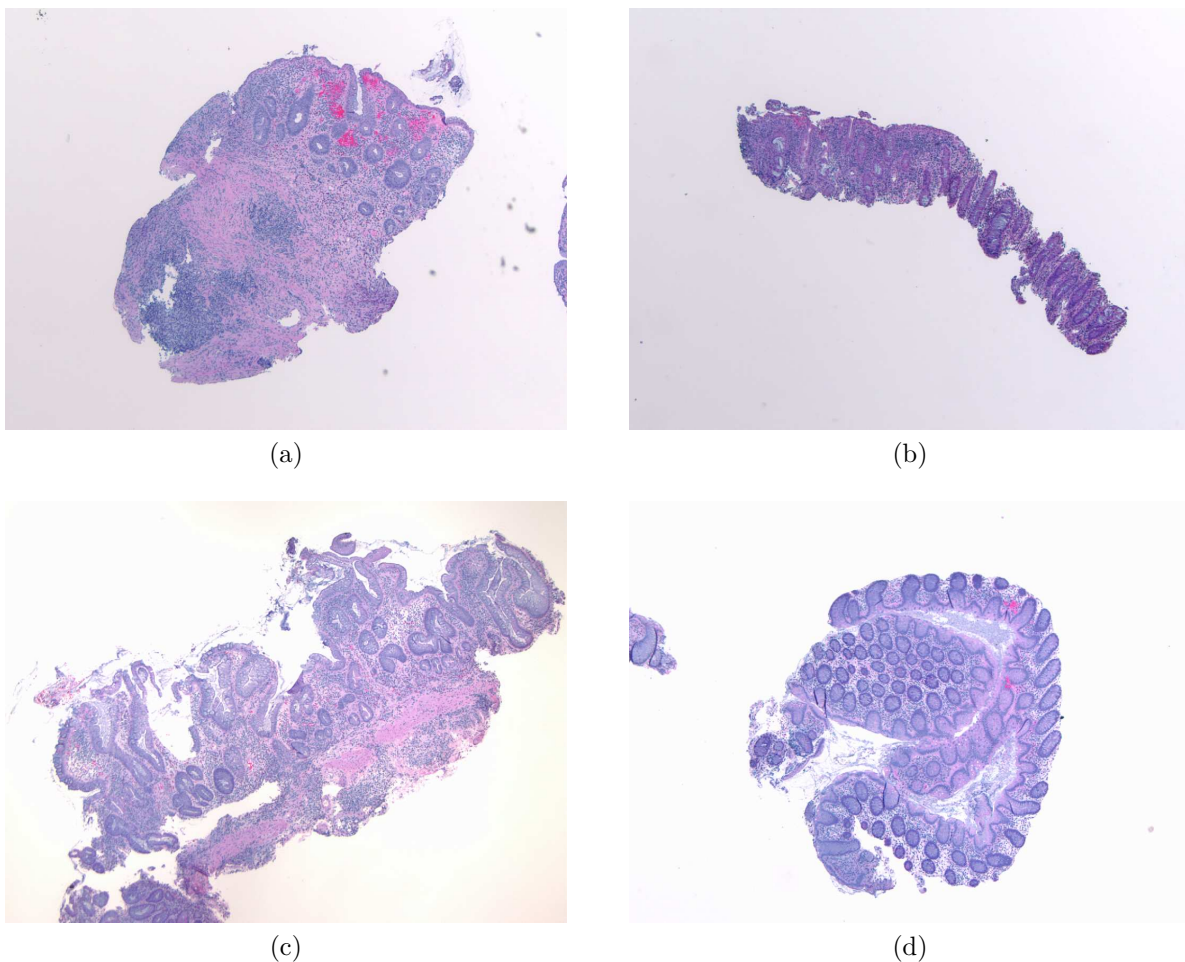
Figure 1.5: Examples of GI tract biopsy variation due to surgery. Each sample was imaged at 40X magnification. Take note of how each sample has a unique shape and how the presentation of some noticeable structures are varied.

The subsequent steps of chemical fixation and histological sectioning are affected by a

myriad of factors. Environmental conditions can greatly affect the chemical reactions that occur during chemical fixation. Additionally, manufacturer to manufacturer variations in the materials can have an impact on this step as well. During histological sectioning, the quality of the sample, the tools, and the technician can lead to gross defects such as missing regions of tissue in the resulting slides.

The final steps of staining and imaging result in more subtle and often more difficult to identify variations. Different staining protocols will lead to different colorations as will the different manufacturing sources of the stains. After staining, the imaging step is influenced by the quality of the optics and digital sensor. High-quality optics are expensive and thus not available in all applications leading to less than optimal image quality. High-resolution CCD sensors are quite common and thus are not a major concern but inevitably there are situations in which they are still insufficient.

Given the many sources of variation and artifacts as a result of the acquisition procedure there are the domain specific challenges to consider. Let us begin with the clinical domain where the primary sources of variation come from the differences in patients and the complexity of diseases that are considered. Gender, age, past and current health are all factors in how any tissue will present itself. Age is very intuitive in that as an individual matures, so do their tissues; immature versus mature tissue can have very different appearances and pathologies. For example, refer to Figure 1.6 to see the difference between immature and mature forms of neuroglial tissue (a type of brain tissue).

Given the variation that arises from a patient's uniqueness, the complexity of the disease or diseases they suffer from gives rise to even more variation. Many diseases are not acute in that they only present one type of change or pathology. These diseases also mature and as a result have a continuum of changes from the less severe at onset to the most severe towards the end of their progression. For example, consider cancer in general where the initial malignancy may be relatively small and not very aggressive. However, as the cancer

(a)   (b)
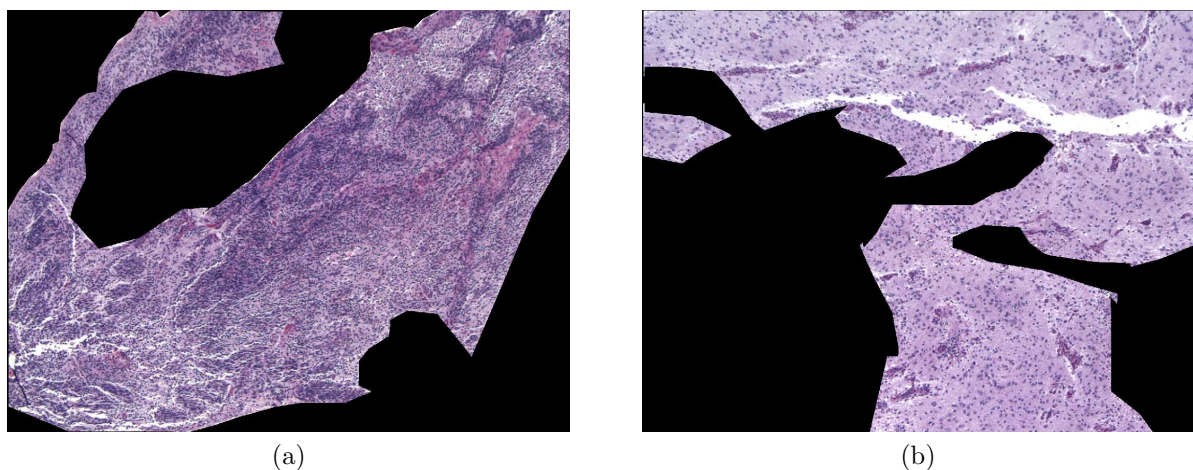
Figure 1.6: Comparison between (a) immature neuroglial tissue and (b) mature neuroglial tissue. Non-neuroglial tissue is masked out in black. The difference in appearance reflects a strong change in tissue structure and organization as a result of maturation.

continues to grow, it may become more malignant and aggressive. These changes are often reflected in the diagnosis as the cancer becomes more severe but as has been stated earlier, clear degrees of severity are difficult to assign. There are many other factors to consider, but we will summarize by stating that a disease's level of progression and overall complexity can give rise to many types of variation in a given sample. Figure 1.7 summarizes the sources of variation in the clinical domain.

Sources of variation in the research domain are primarily a function of the research task. In general, the source of the data being researched is subject to the same concerns as those in clinical domain. However, in some applications, the environment in which the samples are acquired is very different than those found in clinical applications. A research application we will investigate is concerned with the research of teratomas derived from human and non-human primate embryonic stem (ES) cells. We will go into more detail concerning this topic in later sections, but for now it will suffice to state that all possible tissues can be found in these teratomas in any possible configuration. This means examples of the same tissue will often have very different appearances such as bone shown in Figure 1.8. In the context of the classification task we are pursuing, we refer to this behavior as intra-class variability.
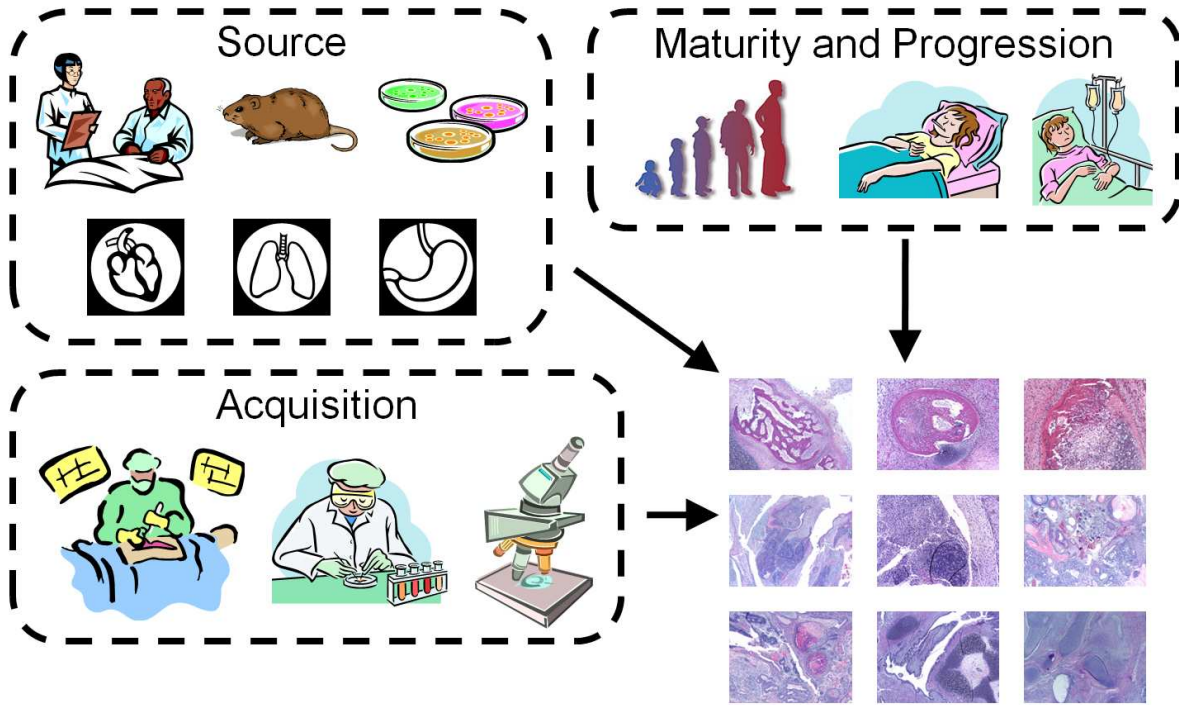
Figure 1.7: Pictorial depiction of the sources of variation that are present in the clinical domain. These sources of variation are also present in the research domain in addition to the other sources of variation that are discussed.

A complementary type of variation is what we refer to as inter-class similarity or situations in which different tissues have similar appearances. The reasons for this behavior range from as simple as tissues being related to each other such as different types of muscle to those so complex that they are beyond our current understanding. Regardless of the reason for these situations, they can be troublesome sources of error in almost any task, especially ones involving the identification of the tissues. Figure 1.9 presents two tissues derived from the teratoma application whose visual appearances are almost indistinguishable to the untrained eye.

The natural variation that arises form the organic nature of the data is the primary source of difficulty as for most applications it cannot be reasonably estimated or predicted given our current level of comprehension. This view also lends credence to the notion that there are more basic and fundamental qualities of tissues that are associated with their identity rather
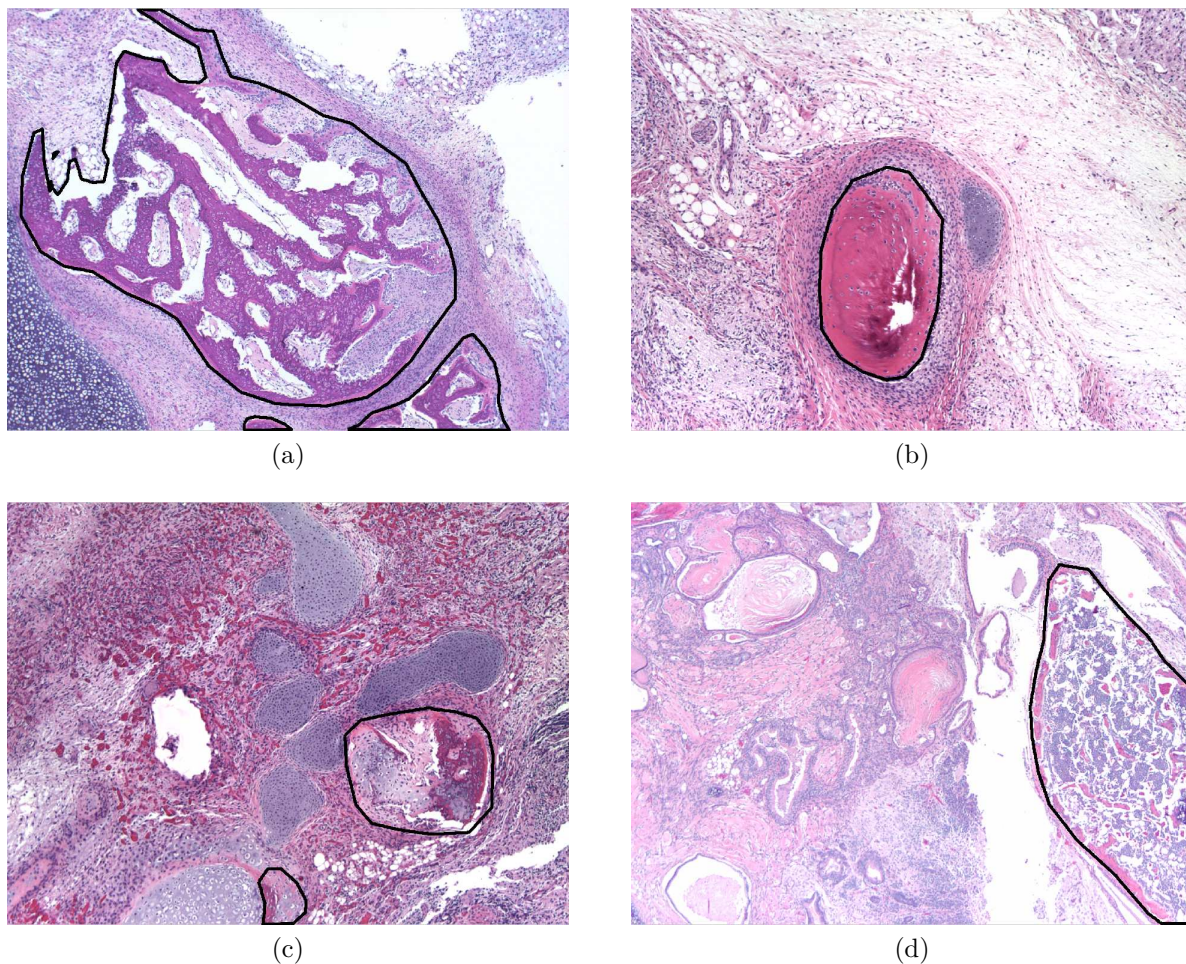
Figure 1.8:   Examples of bone tissue (outlined in black) derived from teratomas. While these are all the same tissue, different environmental conditions along with tissue maturity and configuration have led to very different presentations.

than any one specific presentation. This is a key notion that we will emphasize throughout this work.

1.5.3   Guiding Principles

The importance of histopathology related work along with the associated challenges makes our goal an ambitious and difficult one. In order to accomplish it, we must propose and adhere to principles that encourage success in an informative way from the very inception of the algorithm. In other words, an effective algorithm whose reasons for success are understood
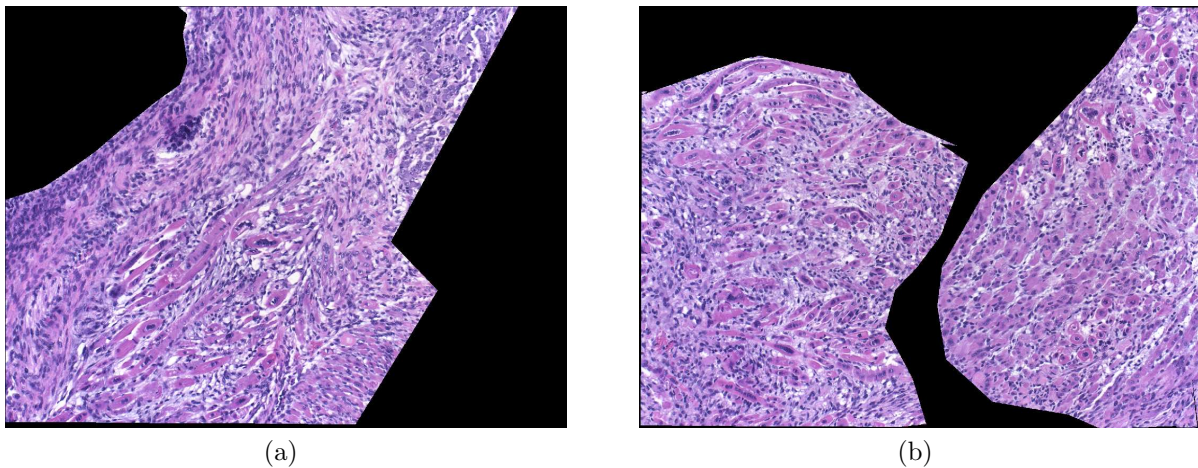
Figure 1.9: Comparison of examples of (a) myenteric plexus and (b) striated muscle tissue both imaged at 10X magnification. Not all examples show such strong similarity but this case underlines the potential difficulty in the task of distinguishing between tissues.

and thus can be further expanded on in later work. This reasoning leads us to pursue our goal according to the following guiding principles:

- *Pixel-level classification.* Rather than assigning a single label to the entire image, we will instead assign a label to each individual pixel based on the features exhibited in the local neighborhood of that pixel.

- *Expert knowledge.* The choice of features we use will be based on the actual visual cues used by pathologists, leading to the development of a *histopathology vocabulary* (HV): a set of common terms understood by both pathologists and engineers.

As we develop our overall methodology and framework, we will refer back to these guiding principles to justify various design choices. They will also help explain various experimental results and our intuition concerning current and future performance. To clarify the overall structure of our classification algorithm, we present Figure 1.10 where the primary computational blocks of our algorithm are shown.
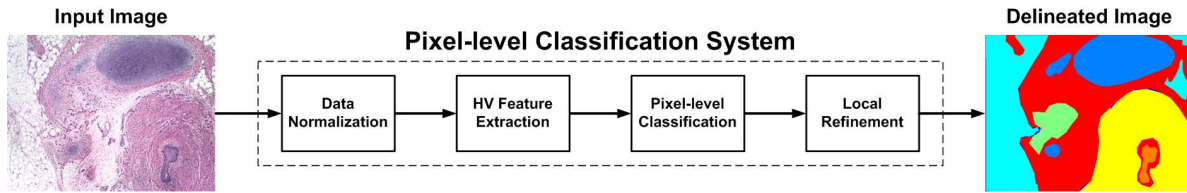
Figure 1.10: Block diagram of our pixel-level classification algorithm. The structure is mostly identical to the standard image classification algorithm with the primary contributions being in the HV feature extraction stage and the additional local refinement stage.

1.6 Contributions of the Thesis

The primary contributions of this thesis are as follow:

- *Mathematical framework for local histograms and occlusion models.* The development of a cohesive mathematical framework that is capable of describing a particular families of images including those relevant to histopathology applications. In particular, the description of complex textures as structured hierarchy of simple textures. This framework is formulated independent of any given application and as a result we expect it to have applicability beyond the histopathology applications we focus on in this thesis.

- *Histopathology vocabulary.* A methodology for the creation of an intuitive feature set that minimizes the gap between existing expert domain knowledge (pathologist) and the design choices (engineer) made in creating features. Not only do these features prove to be effective and concise, but their particular formulation allows us to immediately understand the reason for their success and how to improve them.

- *Pixel-level classification system.* An overall methodology and algorithm for the pixel-level classification of tissues and pathologies in H&E stained images.

- *Experimental results.* Demonstration of the algorithm's performance on clinical and research applications that are very different in scope and goal. The effectiveness of

the single algorithm on both applications provides support to our design choices and overall guiding principles.

Chapter 2

Background

We begin with an introduction to two histopathology applications, the first a research task and the second a clinical one. Following this, we present an introduction to the basics of most modern classification systems that should prove useful to those relatively new to the area.

2.1 Histopathology Applications

As stated earlier, many important histopathology applications and corresponding efforts to automate them exist. Two such applications that we will address in detail in this thesis are first, research related to teratomas derived from ES cells and second, the detection of an inflammation of the colon called active colitis. Both applications are based on analysis of H&E stained images but have many differences in the context of their tasks. By addressing these two differing applications in our work, we wish to demonstrate a commonality amongst them and how to take advantage of it.

2.1.1   Research Application: Teratoma Tissue Delineation

ES cells are pluripotent cells in their capacity to differentiate into any cell in the body and thus form all the basic tissues that may be found. Specifically, they can become cells of any one of the three primary germ layers of mesoderm, endoderm, and ectoderm from which all tissues are derived. They are the fundamental cells that represent the origin of all cells. Furthermore, ES cells have become critical to many exciting areas of biomedical treatment and research for many reasons including their ability to self-renew and perpetuate indefinitely. Tissue regeneration and repair, drug development and therapy, and treatment of developmental and genetic disorders are just a few of the many applications in which ES cells have potential.

In typical laboratory situations, ES cells are characterized by the proteins they express and their behavior in culture. However, a key consideration in human and non-human primates is that cells cannot be considered to be ES cells until they are able to form a teratoma when injected into immunocompromised mice. Teratomas are tumors that are strictly defined by histological evidence of tissues derived from each of the three primary germ layers. Equivalently, a teratoma has no known prior developmental preferences aside from forming tissues belonging to one of the primary germ layers. This is in contrast to more typical tumors in people where the tissues found are strongly related to the tissues found around the anatomical location of tumor.[1] Figure 2.1 shows some examples of H&E stained images of teratomas and their chaotic organization of tissues.

Observation of a teratoma after H&E staining will often reveal masses of many types of tissue with no apparent organization. Furthermore, normal spatial relationships between tissues are not likely to be present. For example, one would not expect a mass of bone embedded in brain tissue but in a teratoma this is possible. From both a research and clinical

---

[1]The term teratoma is not solely limited to the context of research with ES cells. Teratomas have been found naturally in humans and are often congenital conditions (existing from birth). Documented examples of teratomas tumors in patients have been found to contain a wide variety of structures including basic tissues such as skin and bone and in more rare cases, complex organs such as eyes and hands.
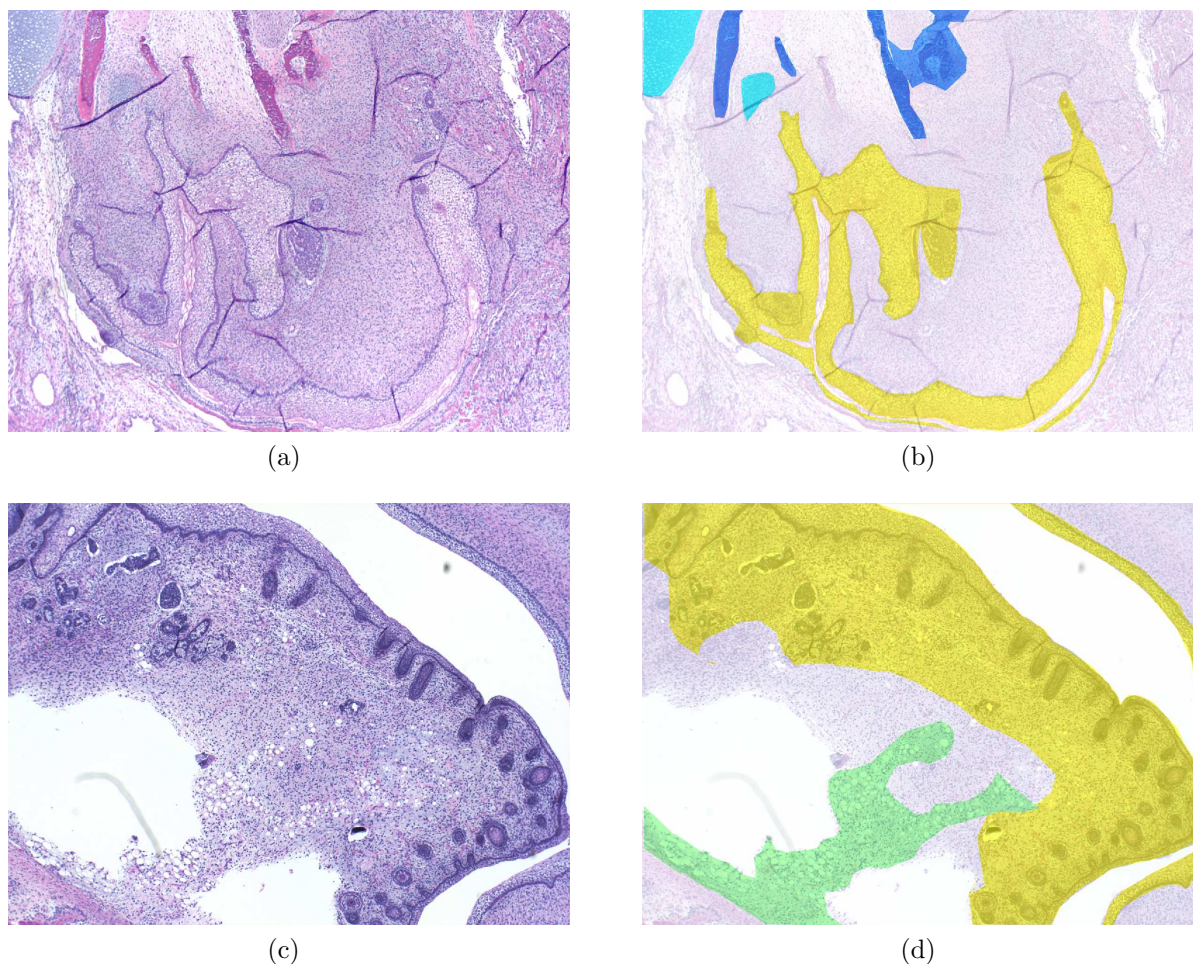
Figure 2.1: Example H&E stained images of teratomas derived from ES cells. The first column presents the original image with the second column showing various tissues transparently colored according to: bone is blue, cartilage is cyan, skin is yellow, fat is green, and gastrointestinal tissue is orange while all other tissues are not highlighted. Note how not all tissues are present in all images and their particular configurations vary greatly. Continued in subsequent figures.

standpoint, a teratoma is seemingly unpredictable in both composition and organization.

While this apparent chaos may be seem unattractive, it is the teratoma's ability to form tissues free of normal developmental constraints that may allow us to answer important questions. Most generally, what are the exact stimuli that cause ES cells to develop normally or abnormally into any given tissue? By understanding these stimuli, we can contribute significantly to many areas. Naturally, understanding the mechanisms of development answers

(e)

(f)

(g)

(h)
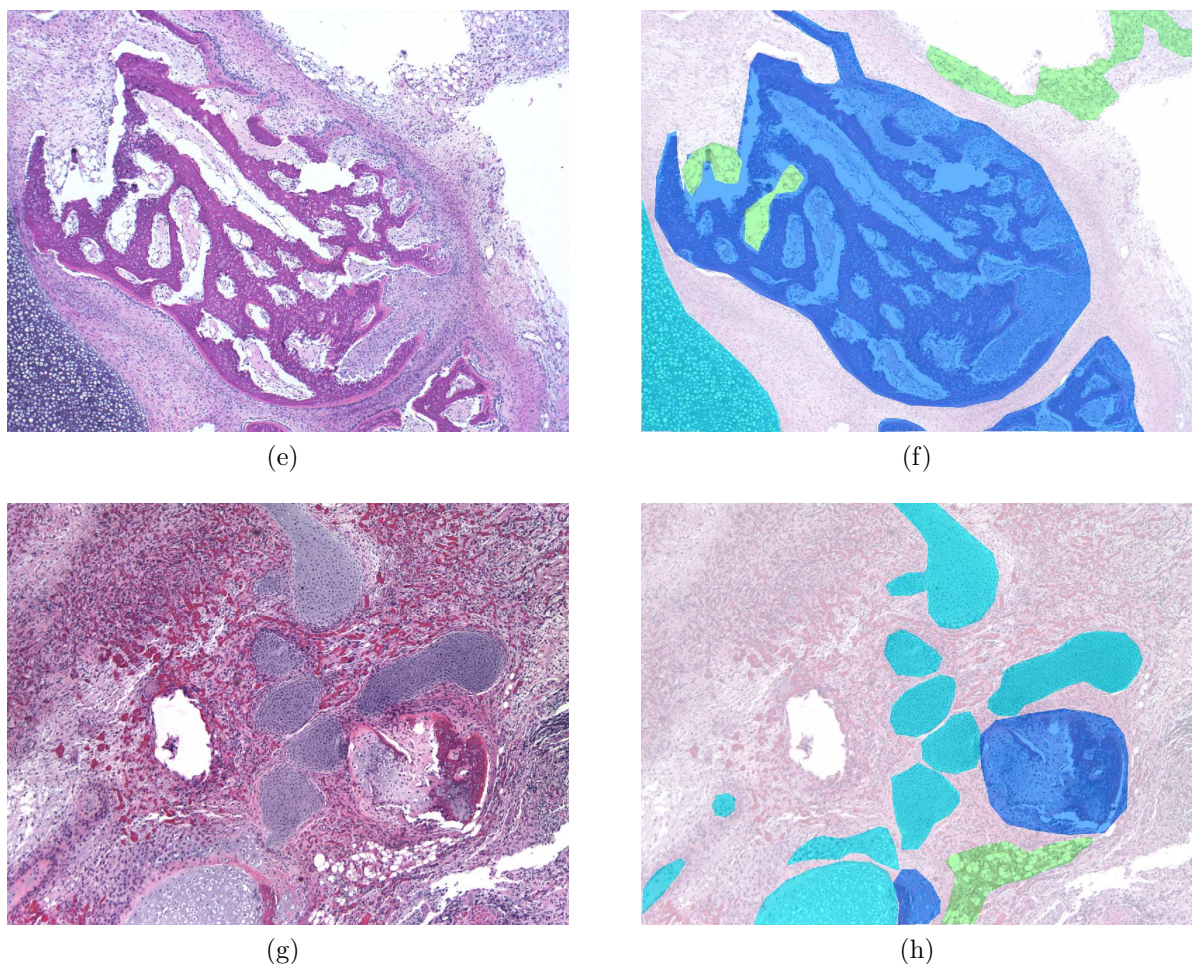
Figure 2.1 continued.

many basic science questions that in turn can greatly aid work in more application oriented areas. For example, in developing tissue regeneration and repair techniques, determining how to cause ES cells to differentiate into particular tissues has clear implications. In particular, understanding how they develop in isolation of normal developmental conditions could lead to methods for culturing of tissue, dramatically increasing the applicability of the methods.

In the context of H&E stained images of teratomas, quantitative knowledge concerning the organization of its tissues is critical to our understanding of normal and abnormal development. This quantitative knowledge can only be extracted after tissues have been identified and delineated. When manually done by pathologists, it is a time-consuming task

(i)

(j)

(k)

(l)
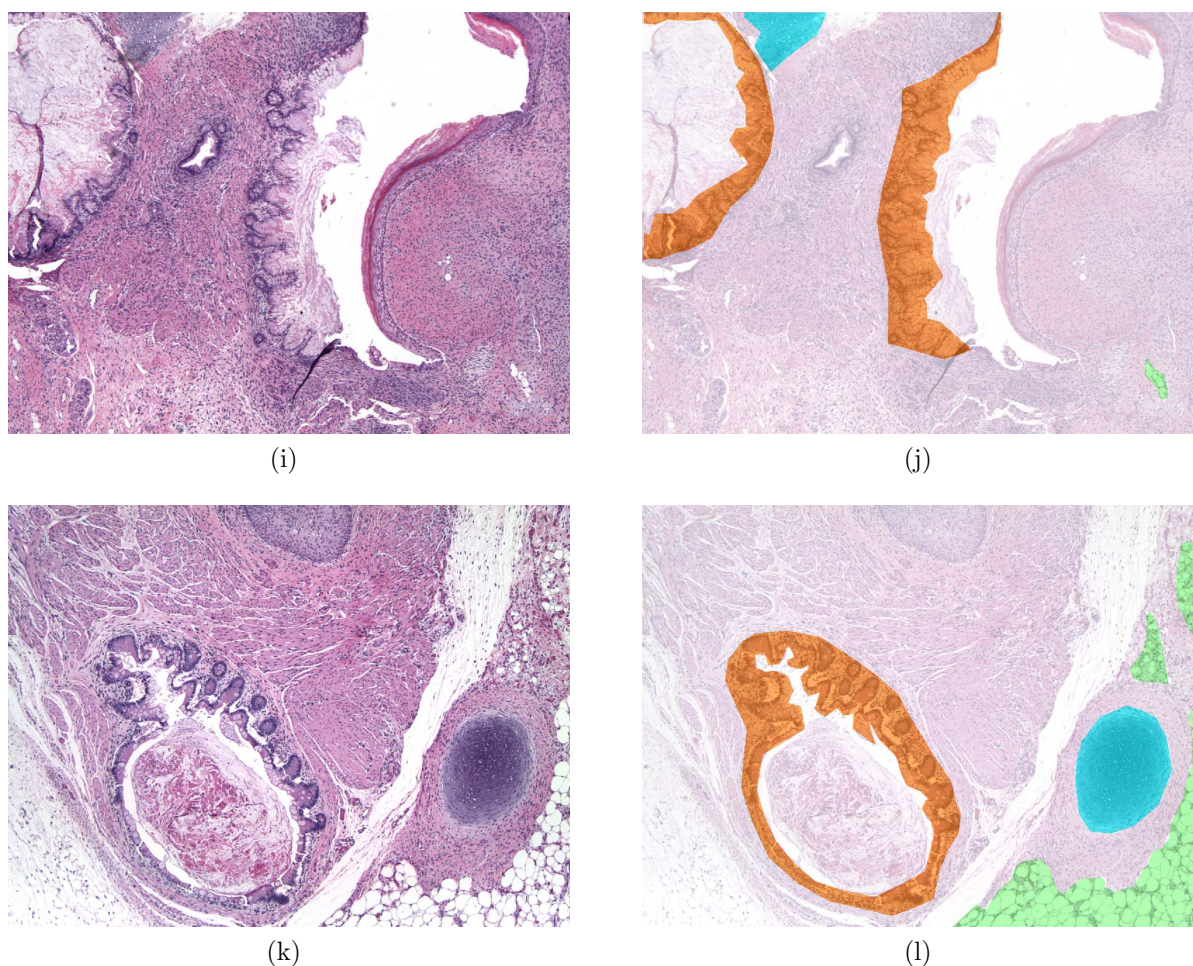
Figure 2.1 continued.

but balanced by their very accurate and relatively consistent results. Such research must analyze vast amounts of data to achieve accurate and useful conclusions. As a result of the commitment of time and the sheer amount of analysis required, the rate of research is somewhat handicapped by the manual nature of analysis. If some level of automation could be introduced, a significant burden of work could be removed from the pathologist and thus increase the pace this research.

However, the extreme variability in the number and configuration of tissues in any given teratoma sample makes this a very difficult task to automate. As stated before, the extremely chaotic nature of the teratoma makes predicting likely configurations of teratomas almost

impossible at this point. In contrast to cancer detection/grading applications where it can be assumed that an image contains only one type of tissue such as breast tissue, this task is a multi-class problem that requires segmentation of any given image into unique regions-of-interest (ROI). Each segmented ROI should contain only one tissue whose identity can then be determined. Unlike classification of tissues in normal human systems, teratomas do not have any known regular organization that can be taken advantage of. For example, in the GI tract, one does not expect to see brain or lung tissue amongst colon tissue though this might be the case in some teratoma samples. Without this context, no expectations can be placed on the exact presentation of any given teratoma sample. As a result of these considerations, this task can be considered as a rather general classification problem.

The overall goal of this thesis is derived from this application since it represents a generalization of many types of H&E stained image analysis problems. Specifically, the requirements for an algorithm to perform well with teratomas include the ability to distinguish between many types of tissues and pathologies (mostly related to maturity), to spatially localize these tissues, and to operate with little or no prior context. This in contrast to other more targeted applications where only a limited number of tissues and pathologies are considered, localization may not be necessary, and there is often a strong context that can be used to simplify the problem. As a result, we will make constant reference to this application in both development of our algorithm and experimental results.

2.1.2 Clinical Application: Active Colitis Detection

Our second application concerns the diagnosis of active or ulcerative colitis, an inflammation of the colon portion of the gastrointestinal (GI) tract. A chronic condition, there is no cure for active colitis though it can be treated with medicine and is often asymptomatic. Among the symptoms of this disease are mild stomach pains, blood and other fluids in waste material, and occasional mild fevers. Methods of diagnosis include endoscopic examination,

X-ray examination, and various types of lab analysis such as blood work. Milder cases of the condition can be treated with medication though more severe cases can require hospitalization and/or surgery to remove portions of the colon. Untreated, persistent active colitis can lead to increased risks of colon cancer.

The number of people affected by or prevalence of active colitis in the U.S. during 1998 was reported to be approximately 619 thousand individuals [26]. According to [27], in the U.S. in 2004 approximately 716 thousand patients were seen by office-based physicians, hospital outpatient clinics, and emergency departments for active colitis. The same study [27] also reported that about 2.1 million active colitis treatment prescriptions were made resulting in a total retail cost of approximately $272 million in the U.S. during 2004. While active colitis is not nearly the most prevalent digestive disease, its impact is considerable in terms of people affected and associated expenses.

As is the case with many diseases, early detection is an important factor for successful treatment of active colitis. Furthermore, accurate diagnosis can ensure that the appropriate course of treatment is taken. The most definitive method of diagnosis of active colitis is an visual inspection and biopsy of the colon facilitated by the use of an endoscope. To better understand the nature of this process, we will briefly discuss some colon anatomy and some specifics of the endoscopic biopsy process.

As shown in Figure 2.2(a), the colon is the penultimate portion of the GI tract terminating in the rectum. Its primary function is the formation of solid waste, maintenance of water balance, and absorption of some vitamins. The colon is effectively a hollow tube whose thickness is formed by a series of different tissues. As a result, the inner lining of the colon (from which the biopsy is taken) is a textured surface. Specifically, as shown in Figure 2.2(b), this texture is the result of the folding of the epithelium of the colon.

The gaps or valleys between the folds of the epithelial layer are very characteristic of the colon and are called glands of Lieberkühn or crypts whose structure is a key indicator of
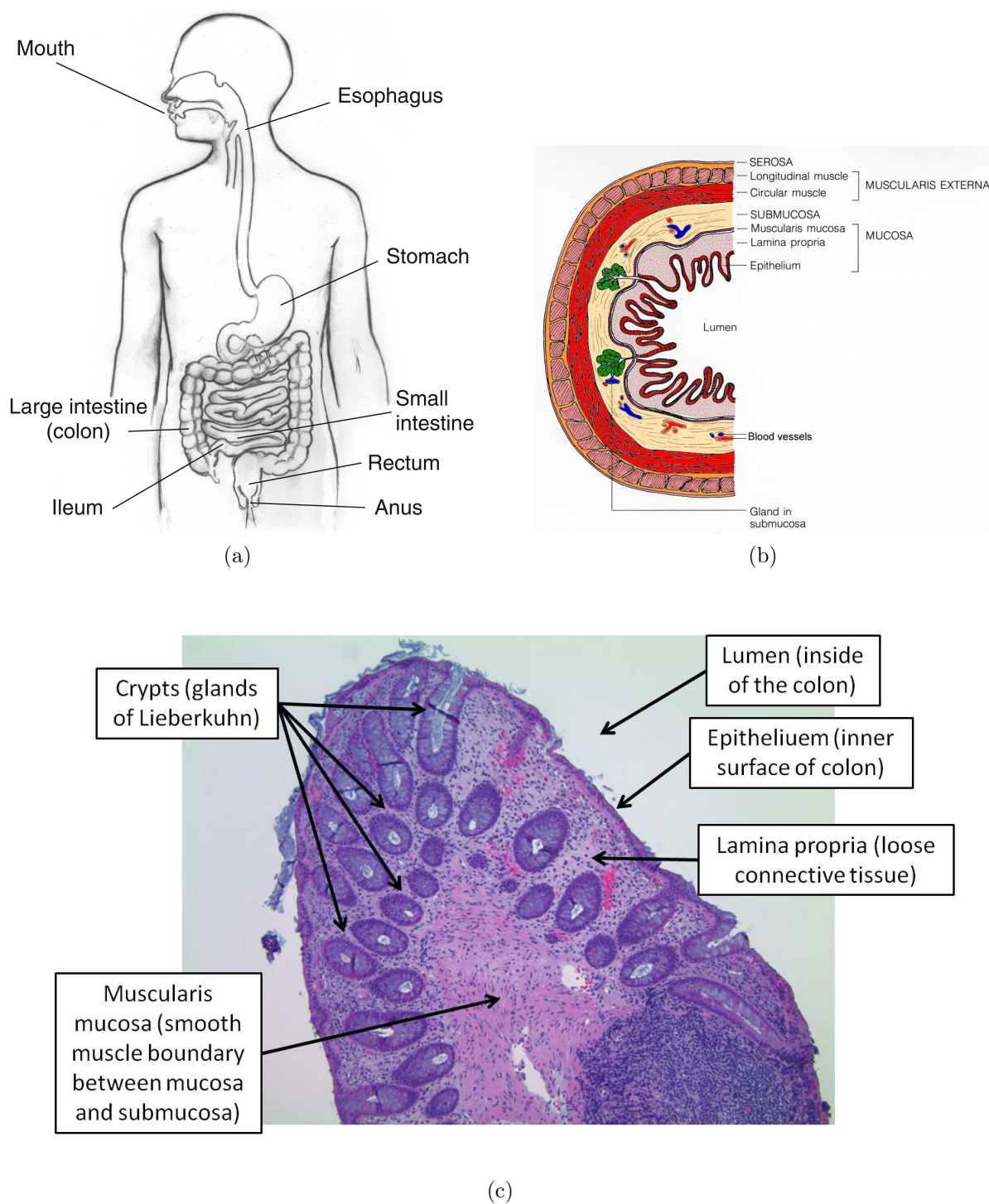
Figure 2.2: (a) Anatomical location (www2.niddk.nih.gov) and (b) transverse cross-section of colon (www.cancerquest.org), and (c) example H&E stained endoscopic biopsy with key components shown.

the presence and severity of active colitis. One can consider the crypts to be like irregular cylinders or cones whose walls are formed by the epithelium. Below this epithelial layer is a sheet of lamina propia, a type of loose connective tissue. Next is a layer of muscularis mucosa which is a type of smooth muscle. The final layer relevant to our discussion is the submucosa which is a dense connective tissue. These layers collectively are the mucosa which is the source of the majority of endoscopic biopsy sample. Figure 2.2(c) gives an example of the presentation of these various components in a H&E stained image of an endoscopic biopsy.

A modern endoscope is effectively a long flexible tube with an illumination source, a camera, and one or more instrument channels at one end and an eyepiece at the other. The end having the illumination source and camera is inserted into a natural opening of the body and guided through the organ of interest which are the rectum and colon respectively in the case of active colitis. The combination of illumination source, camera, and eyepiece allows a physician to visually inspect the interior of the colon for any signs of active colitis. Via the instrument channels, the operator is able to deliver surgical tools (e.g., a forceps and snare) allowing them to acquire a small biopsy of the colon's inner lining. Depending on the nature of the biopsy, different tools and procedures may be used. Figure 2.3 presents an illustrative example of such a surgical tool and the corresponding acquisition procedure.

As one can tell from Figure 2.3(b), the exact location and orientation of the forceps relative to the inner lining of the colon is the deciding factor in the exact presentation of many important structures. As mentioned earlier, the crypts of the colon are such important structures and their appearance in H&E stained images can vary greatly based on the orientation of the acquired sample.

Specifically, imagine cutting a cylinder of some thickness whose ends were closed off along one direction and then observing it perpendicular to the plane of that cut. If you were to cut it perpendicular to its length (i.e., transversely) and then observe it, you would see a
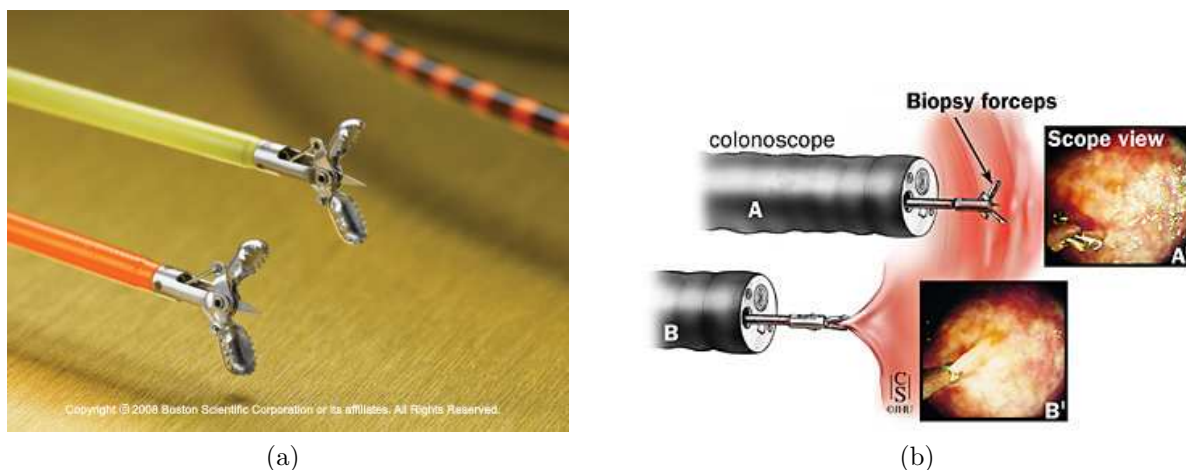
Figure 2.3: (a) Endoscopic forceps (www.bostonscientific.com) and (b) illustrative example of biopsy acquisition using forceps (www.hopkins-gi.org).

perfectly circular ring whose thickness was that of the cylinder. Now consider making that cut along its length (i.e., longitudinally); you would then observe a rectangle whose sides would be that of the length of the cylinder and its ends. Obviously, these two observations are completely different as a result of the manner in which the cut was made. Any particular cut will result in a particular size and shape though they may not necessarily be unique with respect to other cuts. This basic notion applies to the crypts and their presentation in H&E stained images as demonstrated in Figure 2.4.

This source of variation in a given image's appearance can be a major source of difficulty in designing automated analysis tools. Although a pathologist is not likely to have issues recognizing the particular orientation of the biopsy and the structure of the crypts, compensating for the variations that may occur is a challenging task. Based on all of these factors involved in creating a H&E stained image of an endoscopic biopsy, there are some highly indicative cues that pathologists use to discriminate between a colon free of active colitis and one suffering from it. The primary indicators are summarized in Table 2.1.

The level of cellularity in the lamina propia is very indicative of the presence and severity of active colitis. More severe cases of active colitis will present samples where the number of visible cells is dramatically more than it would be normally. The presented structure

(a) (b)
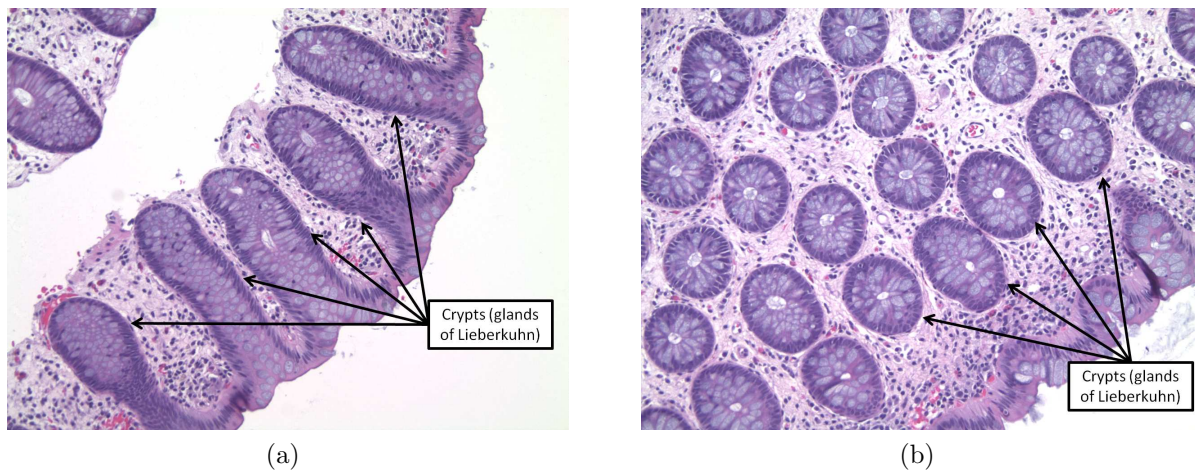
Figure 2.4: Characteristic examples of variation in crypt appearance in H&E stained, 200X magnification images of endoscopic biopsies due to acquisition. Crypt appearance as a result of primarily (a) longitudinal acquisition and (b) transverse acquisition.

Indicator	Normal	Active colitis
Degree of cellularity	Low	High
Crypt structure	Elliptical and well defined	Distorted and fragmented
Crypt organization	Regular	Random
Presence of neutrophils	None	Present

Table 2.1: Key indicators of active colitis and a comparison between them in normal and active colitis affected cases.

(transverse view) of a crypt is a well defined ellipse if not a near circle. However, the onset of active colitis will cause this structure to deviate from an ellipse and often become distorted and fragmented if not completely destroyed. In a similar comparison, the organization of crypts is fairly regular in normal colons while becoming increasingly random in more and more severe cases of active colitis. The final key indicator of active colitis is the presence of neutrophils or white blood cells whose presence is a key indicator of the immune system's response. In colons affected by active colitis, neutrophils can be found the crypt structures though they can sometimes be mistaken for simple blood and vice versa. Figure 2.5 presents some visual examples of these indicators in the cases of a normal colon and one affected by active colitis.
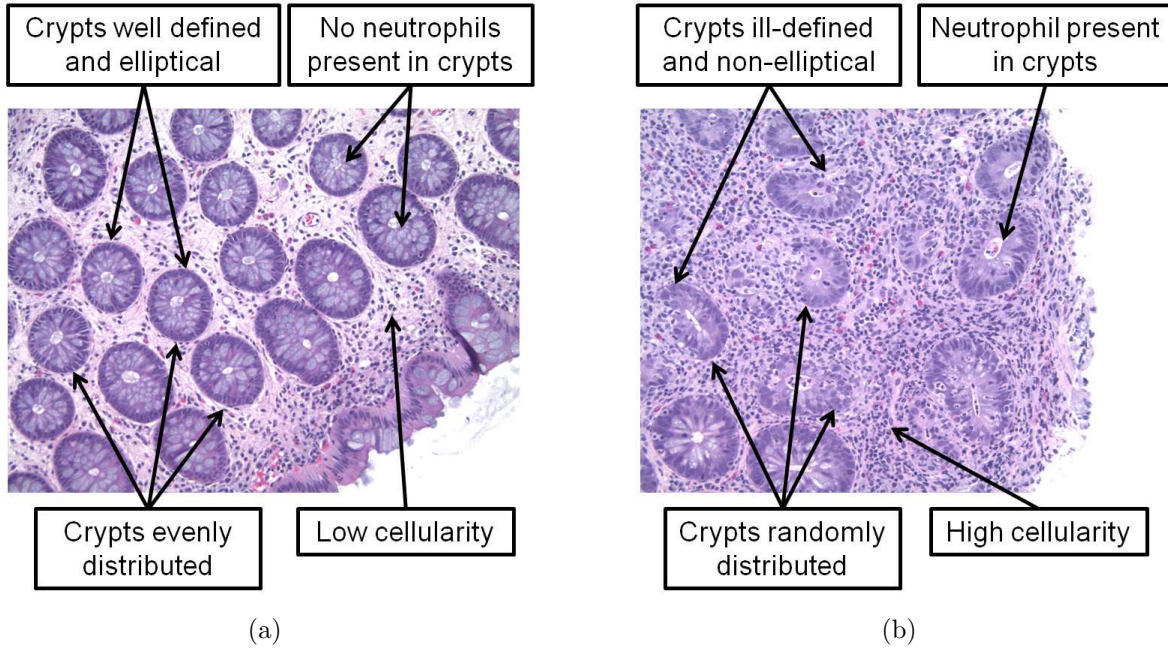
Figure 2.5: Comparison of key indicators of active colitis in a (a) normal colon biopsy and (b) active colitis infected colon biopsy.

To summarize our discussion concerning the colon and active colitis, we present a series of example H&E stained images of endoscopic biopsies imaged under multiple magnifications. Figure 2.6 presents normal or undiseased colons while Figure 2.7 presents active colitis affected or diseased colons. The multiple magnifications demonstrate the distinctness of each condition across different scales.

Similarly to our research application concerning teratomas, automated analysis of these biopsies can greatly improve the overall quality of care in terms of effectiveness and cost. However, unlike teratomas, such analysis is merely a diagnostic aid as opposed to a complete replacement of the pathologist in some role. As stated before, in the current regulatory environment and society's perception of these technologies, no such analysis can assign the final diagnosis and as result, the chosen treatment. Still, such technology can increase a pathologist's accuracy in detecting and correctly evaluating active colitis while also reducing their workload. Some possible uses of such tools were discussed in Section 1.3. Our intent at this time in this application is to utilize our algorithm to perform active colitis detection

(a)

(b)

(c)

(d)
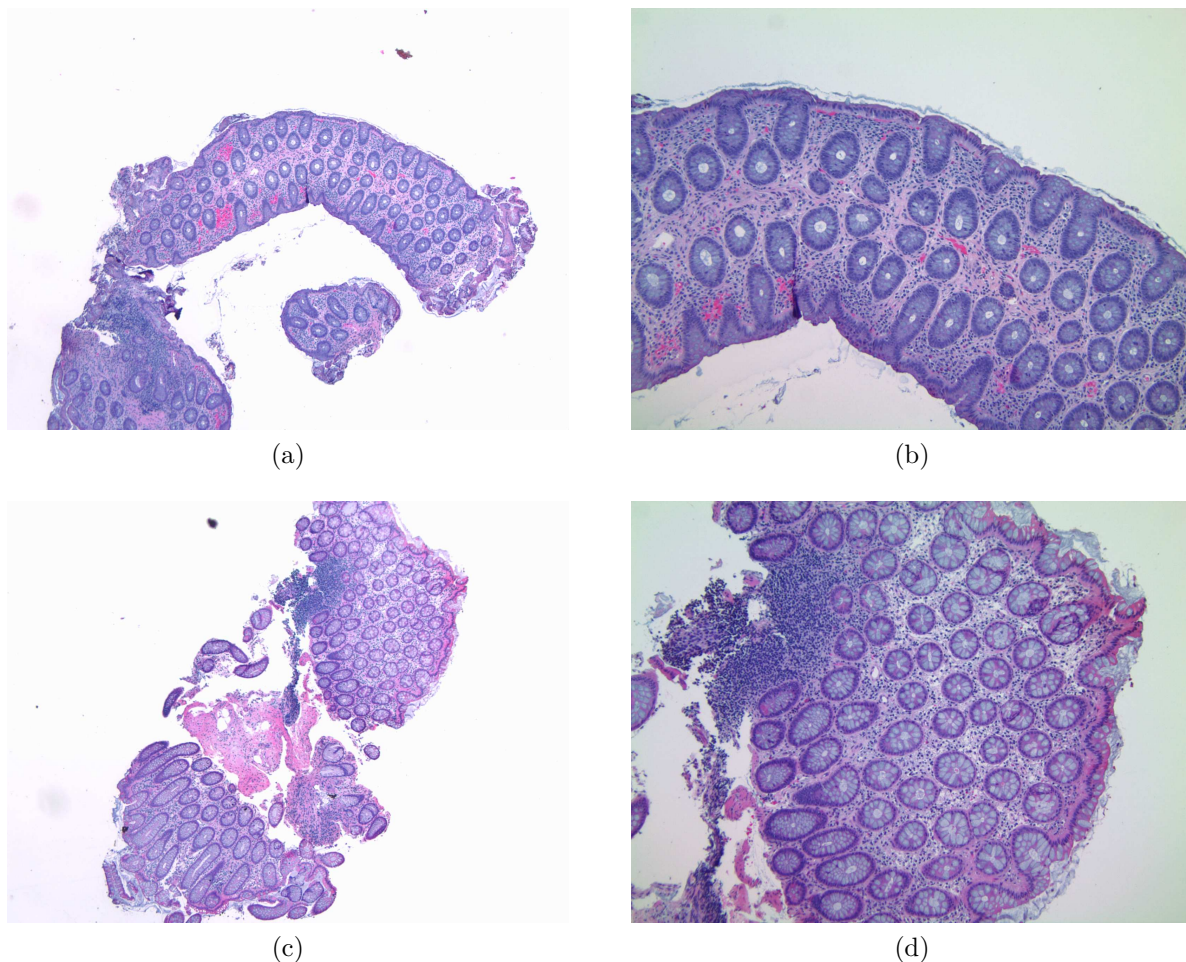
Figure 2.6: Example H&E stained images of endoscopic colon biopsies free of active colitis under different magnifications. Each row is the same sample with the first and second columns presenting it under 40X and 100X magnifications respectively.

in H&E stained images of colon biopsies as a CAD tool.

Unlike the teratoma research application discussed earlier, this application presents a very clear context in terms of how to formulate and solve the problem at hand. To begin with, we are only concerned with images of the colon and differentiating between normal (undiseased) and diseased cases. Furthermore, we can leverage a well established set of image features that pathologists use to evaluate for active colitis. This is in well in line with our guiding principle concerning the design of a HV that reflects both the pathologist's and engineer's knowledge.

(e)

(f)

(g)

(h)
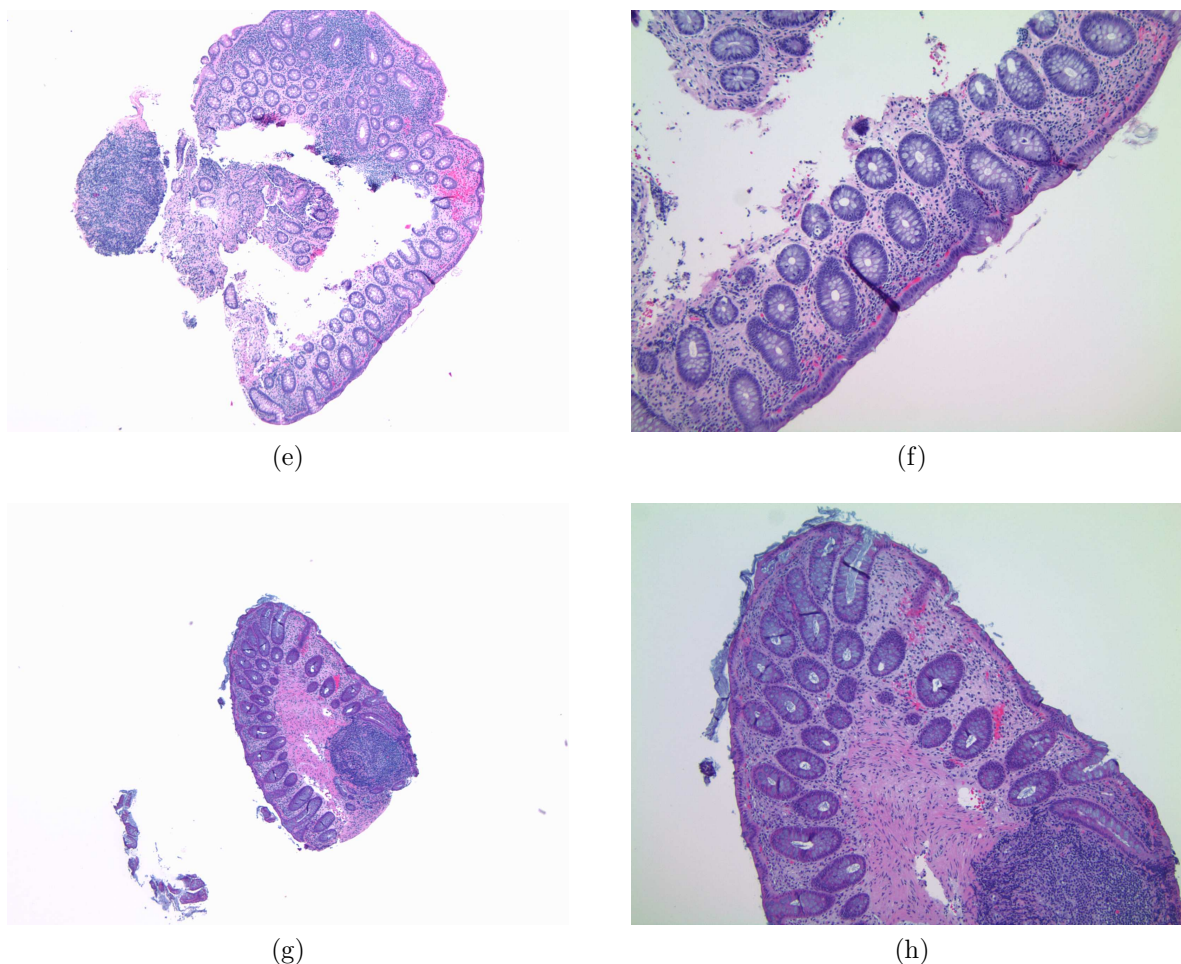
Figure 2.6 continued.

2.2 Classification Systems

The term classification has many meanings but in general it refers to the labeling of objects in a manner that follows some known set of guidelines. For example, the everyday action of identifying objects we see is a classification task. The labels we assign are merely our names for various sights such as a person, a car, a building, and so on. Our single guideline in this task is that the label we assign is correct to the best of our knowledge. The apparent simplicity of this task in its definition and the ease with which we accomplish it belies its actual and extreme complexity.

(a)

(b)

(c)

(d)
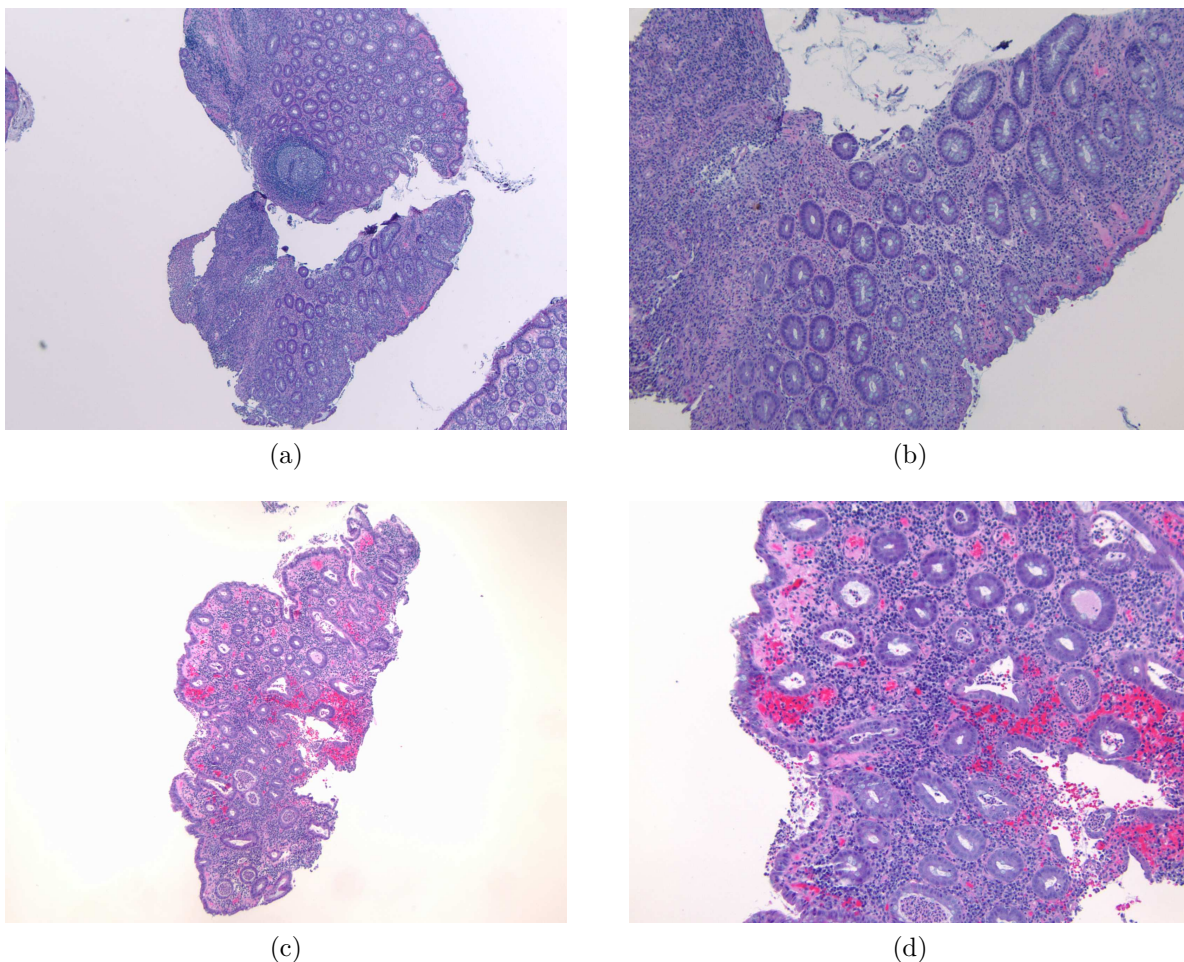
Figure 2.7: Example H&E stained images of endoscopic colon biopsies affected by active colitis under different magnifications. Each row is the same sample with the first and second columns presenting it under 40X and 100X magnifications respectively.

Motivation

With the advent of the digital age in terms of highly capable sensors and computing resources, once manual, i.e., human performed tasks are now being relegated to complex computational systems. The most common goal of such systems is the complete automation of these tasks so as to reduce human burden and improve overall efficiency. It is no surprise then that classification tasks are among such efforts; areas such as robotics, biometrics, automatic target recognition for military applications, bioinformatics, and financial analysis are just a few of the many important fields that are pursuing such computational systems. We refer

(e)

(f)

(g)

(h)
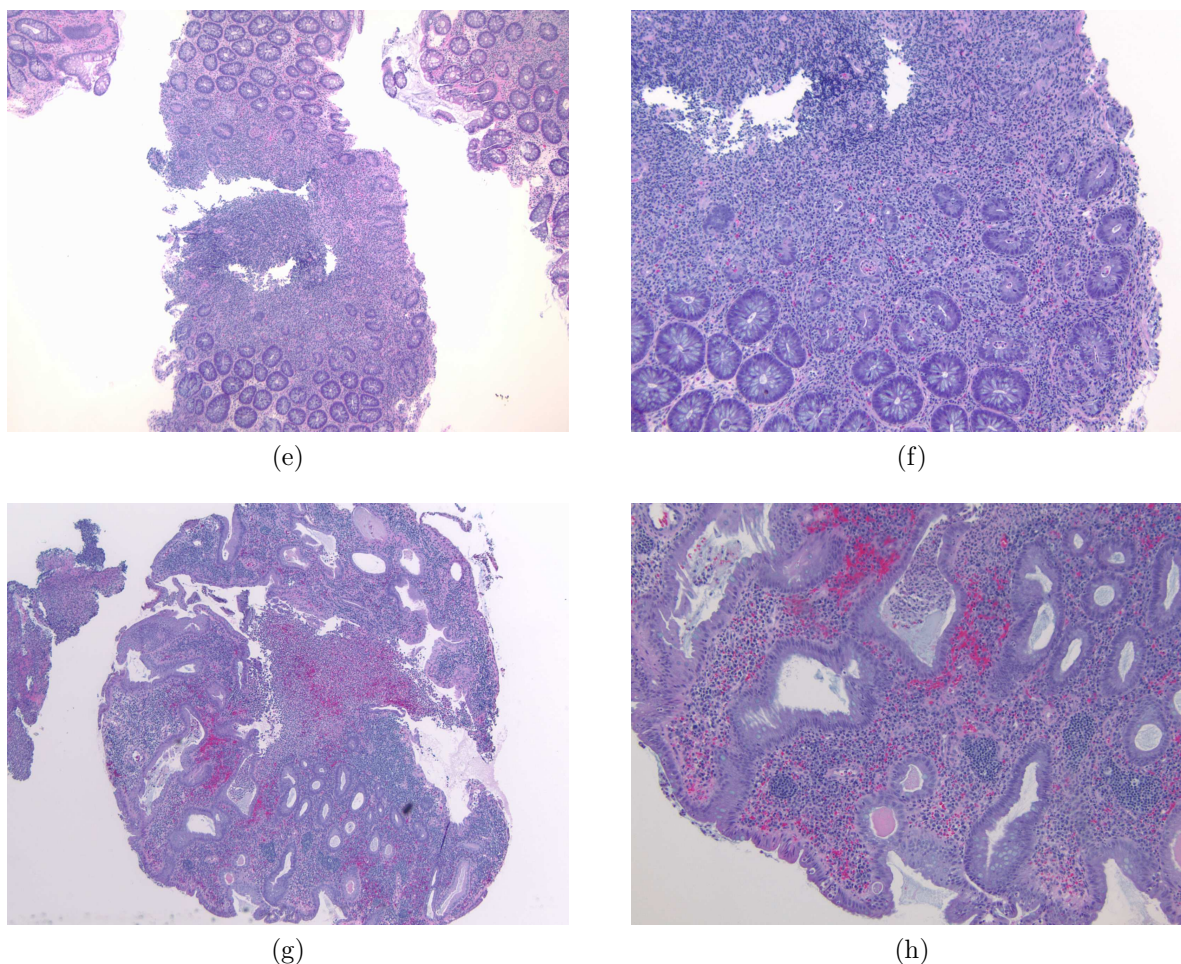
Figure 2.7 continued.

to such computational systems for the automation of classification tasks as classification systems.

Among these many different efforts, the automation of classification tasks involving imagery of some sort is recognized as an important source of technological innovation. As humans are visual creatures by design, there are a myriad of visual tasks (from simple to complex) for which a classification system can be designed. The previous example concerning everyday object recognition is a large and rich area of research that is actively pursued in both industry and academia. Naturally, in this thesis we are concerned with biomedical images in the form of H&E stained images for histopathology.

Design

Given the motivation for creating classification system, their exact designs are strictly dependent on the particular task and application they seek to address. However, the majority of modern classification systems are derived from a prototype architecture involving two main components: feature extraction and a classifier. We will discuss each of these components in turn but first expand on this general architecture.

Figure 2.8(a) presents a simple block diagram of the prototypical classification system. As can be seen, the input to the system is a piece of data which is an image of some sort in our discussion. From the image, descriptive features are extracted using some collection of methods. The goal of feature extraction is to concisely describe a given family of images in a way that promotes the differences between images of different labels and the similarity between images of the same label. For example, if the task is simply to determine if a red-green-blue (RGB) color image is more red than blue, an appropriate feature would be the ratio between the average value of red and blue channels of the image.

These features are then input into a classifier whose purpose is to provide a label for the input image based on the features. A variety of basic classifiers exist with many variations of each available ranging from simple to complex. Using the previous example of determining if an image is more red than blue and using the hypothetical feature, a simple classifier based on a threshold could be used. If our goal is merely to say that the image is on average more red than blue and vice versa, based on the feature, an image could be said to be red if the ratio is greater than 1 and blue if it less than 1. Based on the features used and the nature of the data, some classifiers may be more or less effective.

While this basic system organization is fairly common in much previous and existing work, many variations exist with significant differences. Some systems do not use an explicit feature extraction stage while others may create specialized classifiers that incorporate feature extraction. Additional stages of pre- and post-processing of both images and assigned
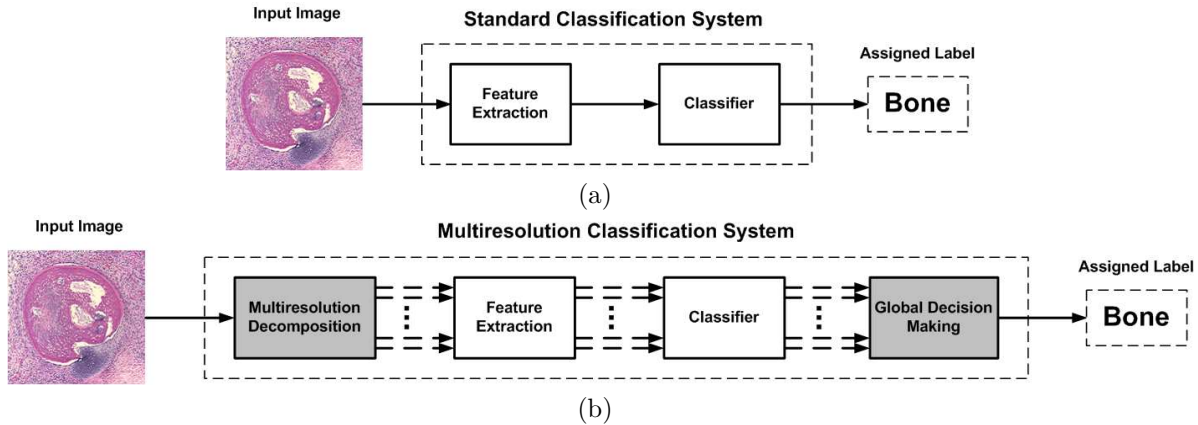
Figure 2.8: Basic block diagrams of (a) standard classification system and (b) multiresolution classification system.

labels are common to reduce errors as a result of noise and other factors. Due to the wide variety of exact designs, our discussion will be fairly general and focus more on items relevant to the contributions of this thesis.

2.2.1 Feature Extraction

The act of describing something to another is very dependent on the language used to communicate between each other. In spoken word, the exact language used has many implications as to how exactly something will be described. It is common that some languages lack clear words to describe certain concepts while others have an abundance of words for some concepts. Furthermore, words that are supposed to have the same meaning in different languages may in fact have meaningful subtleties that are lost in translation. As is typically the case, clear and concise language will transmit the meanings of things much more effectively than one burdened by unnecessary additions.

These same considerations should be taken into account when designing the feature extraction stage of a classification system. Effectively, the features represent the language used to describe the images for the task of classification. Through well-designed features, we can not only ensure accurate and understandable descriptions of a family of images but

also reduce the burden of work on subsequent stages of the system. While there is no distinct image feature research community to speak of, it is a fairly universal concern of those designing classification systems. As a result, the body of existing work is beyond the scope of this work and as such we will only touch base on some general concepts.

One primary motivation in using features as opposed to the original image is to represent data in a more concise manner. Mathematically speaking, this can be considered a form of dimensionality reduction. Any given image is a particular point in a N-dimensional space where N is the total number of pixels in the image. Given the ever-increasing nature of image resolution, N is often very large in most modern image related applications. Furthermore, the volume which any given set of images occupies in this N-dimensional space is typically very small if not infinitesimal. Feature extraction attempts to reduce this N-dimensional space to an M-dimensional space where $M << N$ in order to improve the overall view of the images with respect to the task at hand. As a simple analogy, consider the way you might use binoculars to better see an object that is far away; in a similar manner, the features are zooming in on the characteristics of an image that describe it effectively. However, there are features which in fact do the opposite by increasing the dimensionality to provide the images more space to establish themselves with respect to the task at hand. Even with this consideration in mind, the vast majority of features are dimensionality reducing.

Many image features attempt to describe some intuitive visual quality of the image such as color [28–30], texture [31–34], and shape [35–37]. These features have been researched and developed with great success for a variety of reasons including the strong parallels between them and the manner in which humans perform visual task. Other features may be more complex in describing organization and architecture of certain characteristic components of image such as the positions of the eyes, nose, mouth, and ears on a human face.

Other features are less intuitive in that they leverage more mathematically or statistically oriented descriptions. For example, the Fourier Transform (FT) [38] and other related ideas

have provided a way to describe signals in turns of their frequency content. Such descriptions are often very powerful in characterizing the similarities and differences between samples. Simple statistics such as averages, moments, and histograms are often used to concisely characterize data. In general such features attempt to learn the underlying distributions that govern the data and use them to predict the likelihood that an image belongs to a particular label. A few of the more well known examples of such features include the Viola-Jones object detector [39], the scale-invariant feature transform [40], and image pyramids [41].

Image features are constantly being created and improved upon based on the knowledge each new application provides. Some features are designed to be universal or applicable in any setting while others are specifically designed for a single environment. While it is generally accepted that no one design will solve all problems, for a given family of related applications, it is assumed that a certain set of features will be better suited than others. It is the discovery and design of these features in the context of histopathology applications involving H&E stained images that this work attempts to accomplish in part.

2.2.2 Classifier Design/Selection

Given a fixed set of features, the next step is to design or select a classifier capable of assigning the desired labels to the data based on these features. A classifier is often referred to as a learning algorithm as the goal is to learn a pattern in the data that reflects the specified labels. As a result of this general idea, the terms machine learning and pattern recognition are often used interchangeably to describe much of the related work.

Continuing with our language example from the previous section, if the features represent the vocabulary available to form sentences, the classifier represents the process that parses these sentences and determines their meaning. Given the words in a language, any combination of them can be used to form a sentence but only a subset of them that obey some guidelines are considered proper sentences that have meaning. A classifier's primary

role is to determine this meaning though in some situations it must also distinguish between those cases where there is no meaning.

There are many parallels between research on feature extraction and research on classifier design. Most important amongst them is the large body of existing and continuing work to create new classifier designs and to improve on known ones. Another parallel is that some classifiers are designed independent of the application or problem formulation while others are the opposite in their specificity. As was the case with feature extraction, there is simply too much material concerning classifier design to allow us to provide significant detail in this document. However, there are some general notions that apply to the majority of classifiers that help one to understand where and when a given classifier may or may not be useful.

Training and Testing

The first such notion are the classifier training and testing phases. As the name implies, the classifier training phase is when the classifier learns or trains itself to accomplish the specified task. To facilitate classifier training, some if not all of the available data is designated as the training set while the rest is labeled as the testing set. As the names imply, the training set is used to train the classifier while the testing set is used to test its performance. Occasionally, the data is further divided into a cross-validation set which is still used in the training of the classifier and more specifically to learn or fine tune its various parameters.

Once the training phase is completed, the classifier is considered to be ready for the testing phase. Depending on the situation, the trained classifier is typically presented a never before seen set of data on which it will perform the desired task. In research or proof of concept situations where the goal is demonstrate the potential effectiveness of the classification system, the classifier is applied to the testing set. In a product or deployment situation where the goal is to use the classification system to accomplish the desired task, the data is typically completely new or recently acquired data. Also, in these types of

applications, the new data and the classifier's performance on it can be used to further tune the classifier. Figure 2.9 presents a simple block diagram demonstrating the differences between these two scenarios.
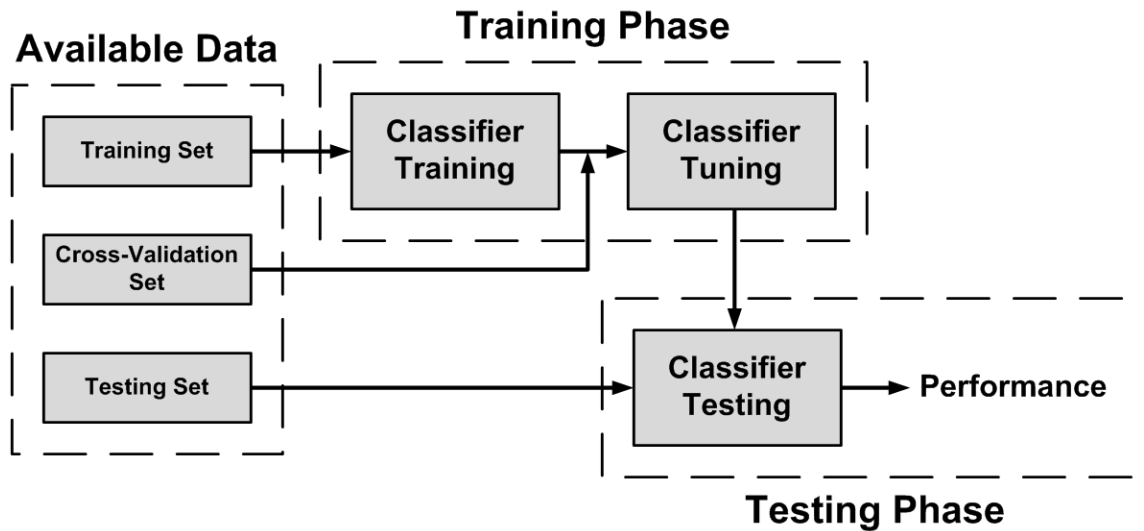
(a)
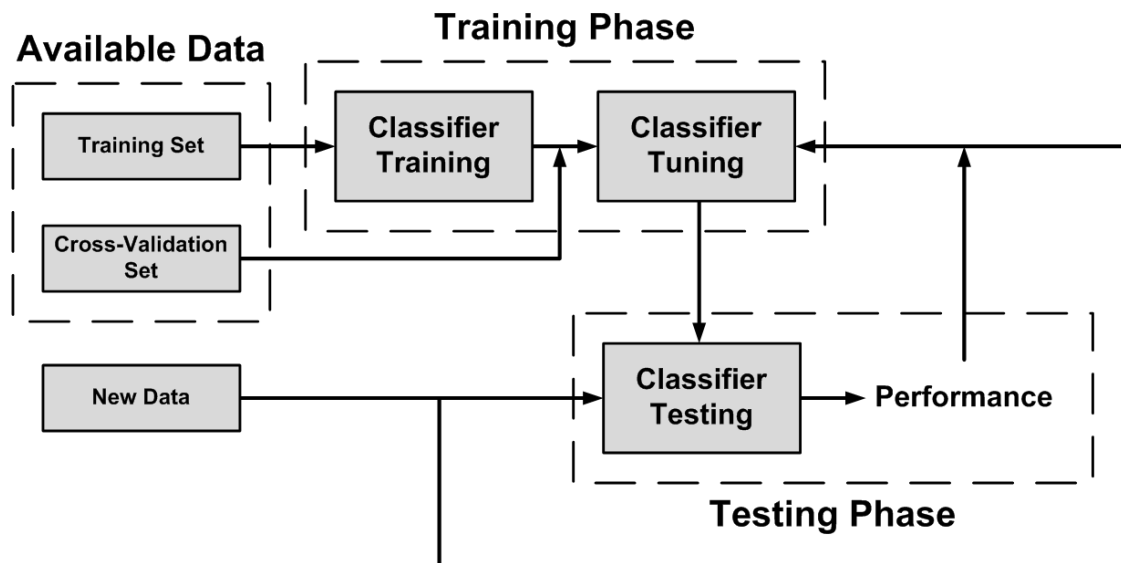
(b)

Figure 2.9: Block diagrams of classifier training and testing phases in (a) research setting and (b) deployment setting.

Supervised, Unsupervised, and Semi-Supervised Learning

Another important concept is that of the differences between supervised, unsupervised, and semi-supervised learning. Simply put, supervised learning refers to situations in which the labels for the training data are known and provided to the classifier during its training phase. This is in contrast to unsupervised learning where no labels for the training data are available. Lastly, semi-supervised learning occurs when some of the training data is labeled while the rest is not. It is often the specific classification task of interest that dictates which of these situations apply. For example, if the task is to determine the identity of a person based on an image of their face, this is a supervised learning situation since we cannot identify a person without teaching the classifier who any given training image corresponds to. Similarly, if the task is to group them together based on similar appearances, this is an unsupervised learning situation since the classifier does not need to know the individual identity of an given image. Semi-supervised examples typically involve some mixture between these examples where the goal may be to determine the identities of samples if they belong to known labels and group them with other unknown labels if they have no such label. Naturally, there are many more examples and variations of these basic ideas.

Dimensionality Reduction/Augmentation

Most classifiers can be categorized as either dimensionality reducing or augmenting methods. The motivation behind dimensionality reducing classifiers is the pursuit of an even more concise description of the images specifically to accomplish the classification task. Most such methods will attempt to determine some subset of the original features and/or some set of combination of the features that maximize performance. A common assumption with such methods is that not all the features are individually useful to the task at hand even if they were designed to be. In contrast, classifiers that augment the dimensionality assume that the existing features are insufficient for the task. As a result, these methods enlarge the

space that the data, as represented by the features, occupies. These new dimensions are often some combination or transformation of the existing features.

Existing Classifier Designs

Based on all these basic notions and more, a large number of classifiers have been developed with many variations on each of them. Among some of the more famous and popular ones are principal component analysis (PCA) [42], linear discriminant analysis (LDA) [43, 44], SVMs [45], neural networks (NN) [46], nonlinear dimensionality reduction [47], and kernel density estimation [48]. These methods have been developed from all manners of standpoints including mathematics, statistics, computer science, and engineering. As a result, some methods have very strong formulations and statements about expected levels of performance for a given task while others are purely algorithmic in the pursuit of computational efficiency and adaptability.

The choice of algorithm for our algorithm is important but not a primary research topic. Rather, we experiment with existing classifiers to determine the optimal one to use in conjunction with our methods. However, in later sections there will be some discussion on classifier designs that may be more appropriate to the family of classification tasks that we wish to address.

2.3 Classification Problem Types

We've discussed the basics of classification systems and specifically those that attempt to automate tasks typically done manually. These classification problems can be grouped according to whether or not a human can accomplish it and if any specialized knowledge is required. Based on these criteria, we define three types of classification problems: *non-expert*, *expert*, and *unknown*.

2.3.1 Non-Expert Classification Problems

Non-expert classification problems are those that a person without any special knowledge can perform. These problems are often those everyday tasks humans perform subconsciously and without any sort of formal training. For example, the most universal non-expert classification problem is everyday object recognition which we now discuss to illustrate aspects of this problem type.

The human ability to accurately identify everyday objects in spite it's huge scope and variation is extraordinary. However, its actual complexity is hidden by how easily humans accomplish it. In general, there is little understanding about the actual mental process behind this task. Since humans are not born being able to recognize all objects, there must be some sort of self-training as we encounter new objects.

In spite of our lack of knowledge about the actual neurological process behind this task, when asked how one recognizes an object, there are some very common answers. Most people will cite descriptive qualities such as color, shape, and size as major indicators of an object's identity. That identity will be unquestioned at some level, e.g., a human is a human though their apparent age may vary from observer to observer. The universality of the descriptive terms and the consistency of object identities implies that, in general, this task does not require specialized learning. Rather, humans are equipped with the tools and some intuitive, though unknown, learning process to accurately and consistently recognize objects. Other examples of non-expert classification problems include facial recognition and scene identification.

For such problems, a person provides an example of a successful system upon which the design of an algorithm can be based. While the design does not need to take any cues from human execution of the task, it would be shortsighted to completely ignore it in light of the accuracy and speed with which it is manually performed. Since an engineer is more than likely capable of the task at hand, they can serve as the working example to learn from as

opposed to requiring input from others. Their knowledge of the task can aid many aspects of classification systems, but feature design often benefits the most from such an approach.

Continuing with object recognition, if the task is to recognize objects at large, an engineer can mimic his own process and use basic features such as shape, size, and color. Similarly, if the task is to recognize a particular category of objects, e.g., cars, then the features might be the presence of wheels, headlights, doors, and windows. In both of these problems, almost anyone is more than capable of proposing some reasonable design choices.

However, it is not always the case that people understand how they perform certain tasks. Some tasks may be so intuitive or subconscious that extracting conscious knowledge of any form may be very difficult. For example, in recognizing colors, most people will not be able to give a clear explanation and instead simply state that a color is a color. However, science provides an alternative understandings of such a fundamental concept (e.g., wavelengths of light) that would allow for the design of a corresponding classification system.

2.3.2 Expert Classification Problems

Expert classification problems are those that only specially trained individual can manually perform making them highly non-intuitive to the average person. For some tasks, the non-intuitive nature is a result of the unique and uncommon modality of the data, e.g., X-ray or MRI. In other tasks, it arises from the task itself such as determining the presence of a brain tumor from a computed axial tomography (CAT) scan of a person's head. In both cases, specialized knowledge (information that must be learned) about some aspect of task is needed.

As a general example, biologists and physicians utilize a number of unique data modalities designed to provide very specific information. Such modalities include X-ray, MRI, CAT, electrocardiography, fluorescence microscopy, DNA sequencing, western blots, and histopathology; each with a unique form and purpose. If you consider the large variety of

modalities and the need for specialized knowledge to interpret them, it is clear to see why various types of biologists and physicians are considered specialists.

Continuing with biology and medicine, the uniqueness of the modalities often lead to highly specialized professions. Radiologists diagnose and treat illnesses by interpreting many types of medical imagery including X-ray, MRI, CAT, and ultrasound. A cell biologist is an expert on fluorescence microscopy and how to use it to understand cell structure and function. While the pairing of modality to specialist is not necessarily one to one, any given individual can only be expert in so many modalities and related tasks.

The specialized knowledge needed to perform these task often makes it difficult to automate them. In contrast to non-expert classification problems, an engineer will not possess the required knowledge, making him ill-equipped to learn from the existing manual solution by himself. The fact that specialists are capable of performing these tasks at significant levels of performance should encourage engineers to leverage their knowledge.

A general analogy is that of learning to play a completely new sport. When you begin, you must learn basics including the rules, equipment, training, and strategies of the sport. Given enough time and examples, you could learn all these things through observation though there is no guarantee you would be proficient. The preferred approach is to learn from an expert, i.e., a trainer or coach, in order to improve your skills. Not only will these experts teach you the fundamentals but also teach you about the more complex aspects of the sport. Similarly, the design of a classification system for expert classification problems can benefit greatly from the knowledge of experts already capable of accomplishing them.

2.3.3 Unknown Classification Problems

Unknown classification problems are those that people are not capable of and for which the appropriate knowledge is unknown. Generally, these tasks represent some novel idea such as a new data modality or a new type of science. For example, given a new data modality,

one of the first tasks is to learn what it is useful for beyond its original purpose. Given the data this new modality generates, understanding of patterns in the data may provide insight about the modality's applicability. Data mining techniques, more specifically, classification systems are one way to extract and quantify these patterns, especially when the data is large and complex.

We wish to emphasize that such unknown classification problems do not provide any proven manual solutions from which to learn. Without such a reference, it is within reason, if not without choice, that an engineer propose completely new methods in his design. These types of problems often represent the limits of human interpretation in a variety of ways such as intractable amounts of data, overwhelming complexity, and a simple inability to interpret the data. Of course as work on a particular unknown classification problem progresses, later works will build upon the successes and failures of earlier ones and eventually change the problem into one of the other two types.

Chapter 3

Previous Work

Before we proceed, a discussion of our previous work in this area is needed to further justify our guiding principles. Most of this work was through the efforts of Chebira et al. [49, 50] utilizing a different algorithm than the one being presented in this thesis. The primary goal of that work was to make an initial effort into automating some aspect of the teratoma research detailed in Section 2.1.1. In this section we will explain the basics of the algorithm developed and also the significance of the results achieved.

3.1   Single-Tissue Classification

Rather than attempt to immediately address the very challenging task of identifying and delineating tissues in teratomas, Chebira et al. chose to demonstrate the feasibility of automating it by first developing an algorithm for single-tissue classification in teratomas. If we were able to achieve reasonable performance with this task, we would have demonstrated that in the context of teratomas, tissues do present enough discriminative information in order to identify them. Given success in this task, subsequent efforts to address the more general task of both identifying and delineating tissues would be better justified.

The fundamental difference between single-tissue and multiple-tissue classification is the

content of a given image. In multiple-tissue classification, an image may contain any number of tissues in any particular organization. The classification task is the delineation task discussed in Section 1.1.4 where tissues must not only be identified but also localized or segmented. This is in contrast to single-tissue classification where the assumption is that an image only contains one type of tissue and the classification task is to identify that type for any given image. This assumption is equivalent to assuming that given a multiple-tissue image, there exists some method (whether manual or automated) to divide the image into single-tissue regions who can then be classified.

Our primary algorithm for single-tissue classification [49, 50] was originally designed as general image classification system. It has been used with success in a variety of applications including high-throughput drosophila embryo screening [51], classification of protein sub-cellular patterns [50], and biometric fingerprint recognition [52]. As the scope of this algorithm is vast, we will explain it in only moderate detail and refer readers to the referenced works for additional information.

3.1.1 Multiresolution Classifier

As discussed in Section 2.2 and shown in Figure 2.8(a), the standard classification system model is primarily a feature extraction stage followed by a classifier stage. This formulation typically considers the input image as the primary representation of the data. The feature extraction stage and in part the classifier stage transform the data to aid the classification task at hand. However, it is our belief that for many families of images, regardless of the classification task, there is a transformation of the data that is beneficial. In other words, a transformation of the data that precedes the feature extraction stage; more specifically, a multiresolution (MR) decomposition.

Figure 2.8(b) presents the overall structure of our MR classifier. As shown, the standard classification system architecture is joined by a MR decomposition stage and a global decision

making stage which we will explain in the following sections in moderate detail. We refer readers to [53] for a detailed explanation of this classification system.

Multiresolution Decomposition

Without going into rigorous theoretical justifications, we proposed to use members of a family of transformations collectively referred to as MR decompositions in this role. These transformations seek to describe data in terms of a series of subspaces. Each of these subspaces represents the data in terms of a particular spatial and frequency support/resolution.

For example, the original signal in the Dirac basis [38] representation where each sample in the signal is a subspace of the smallest spatial support or equivalently the highest spatial resolution. However, this extreme spatial resolution comes at the cost of the frequency resolution since none is provided in the Dirac basis. In contrast, the FT results in a signal where each sample is a subspace that has the highest frequency resolution at the cost of no spatial resolution. Based on the FT, one can tell what frequencies are present in the data but not where they are.

These two transforms illustrate an important concept called the uncertainty principle. This concept states that no single transform is capable of compactly and uniquely representing the data while simultaneously having the highest possible spatial and frequency resolutions. Equivalently, increases in one type of resolution, spatial or frequency, will result in a loss of resolution in the other. As a result, MR analysis transforms are designed to achieve different balances between these two types of resolutions based on the idea that one type of resolution may be more important than the other in certain situations.

Furthermore, MR transforms are often iterative or pyramidal. Unlike the FT, many MR transforms perform a basic set of decompositions on a given signal and the resulting subspaces multiple times. Based on the nature of the decomposition, each iteration may increase one type of resolution and correspondingly decrease another resolution. Additionally, many

transforms seek to provide the ability to reconstruct the signal from the derived subspaces. As a result of these considerations, most MR analysis transforms begin with low frequency and high spatial resolutions with subsequent iterative decompositions increasing and decreasing them respectively.

Transforms such as the short-time Fourier transform (STFT) [38] and the discrete wavelet transform (DWT) [54] are non-adaptive transforms. Such transforms apply iterative decompositions to particular subspaces and their derived subspaces resulting in a structured pyramidal decomposition. Others such as the wavelet packet [54] are adaptive in that they choose which subspaces to apply additional decompositions to based on some metric. These metrics indicate whether or given subspace or the subspaces derived from its decomposition are more useful to the task at hand. The formulation of these metrics is varied depending on the application [55, 56].

A final consideration is the redundancy of the decomposition. Many MR decompositions are non-redundant in that represent the data in as few subspaces as possible. Each subspace is expected to have minimal overlap with other subspaces in terms of their space-frequency coverage. In the interest of efficient representation and compression, non-redundant decompositions have much use such as in the JPEG2000 image format. We refer to a non-redundant decomposition as a multiresolution basis (MRB). Its redundant counterpart is the multiresolution frame (MRF). Rather than describe the data in as few subspaces as possible, subspaces in a MRF have overlapping coverage of the space and frequency. In applications such as robust or error-tolerant coding, such redundancy is useful.

Both MRBs and MRFs are also useful in the context of the MR classifier as they offer efficiency and redundancy respectively. Chebira [53] presented a rigorous analysis of the conditions under which each family of transforms was more or less appropriate. As such justifications are not the focus of this thesis, we refer readers to the referenced work for detailed explanations and proofs of these concepts.

In general, MR decompositions yield a series of subspaces that are expected to collectively hold information better suited to the classification task than the original image itself. Furthermore, the iterative nature of many MR decompositions and the adaptive nature of others can create configurations of subspaces specific to the task. Therefore we expect the use of these decompositions to be generally applicable regardless of the specific classification task.

Feature Extraction and Classifier

The applied MR decomposition yields a series of subspaces each of which represents a particular portion of the space-frequency space. From each of the subspaces we extract a specified set of image features expecting that the values extracted from certain subspaces will be more informative than others. The choice of features is only restricted to those that are computed over an entire image disregarding image resolution. During the training phase, each training image is decomposed into these subspaces with features being extracted from each of them. As a result, we have a set of training feature vectors for each subspace based on the training data which can be used to train classifiers for each subspace.

The intuition behind training an individual classifier for each subspace is similar to a voting strategy. Each subspace represents a unique view of the data and given an appropriately trained classifier, can make a decision about an image based on the corresponding subspace of that image. With many subspaces, many subspace specific decisions are made under the assumption that some are generally more accurate and reliable than others. The same classifier is trained in each subspace with the specific choice up to the designer.

It is this subspace specific nature of the feature extraction and classifier stages that reflects the MR aspect of the algorithm. The parallel arrows between the discrete stages in Figure 2.8(b) represent the subspaces derived from the MR decomposition. At this point in the algorithm we have derived many subspace specific decisions that must now be resolved

into one global decision concerning the input image.

Global Decision Making

This stage of the algorithm combines the previously computed subspace specific decisions into a single global decision for the input image. Many options exist to accomplish this task including a simple voting scheme where the majority decision is assigned as the final global one. However, such an approach would ignore the intuition concerning MR decomposition that certain subspaces contain more useful information than others. To follow this intuition, two different approaches were developed and experimented with. Both methods are simple weighted linear combinations of the decisions made by each individual subspace based on the intuition that more accurate and reliable subspaces will be weighted more heavily. The difference between them is the manner in which they learn the weights for this combination. We use a cross-validation set that has not yet been seen by the classifier to determine these optimal weights.

The first method begins by assigning equal weights to all subspaces and then iteratively updating them. Each sample from the cross-validation set is sequentially decomposed according to our specified MR transform from whose resulting subspace features are extracted and then evaluated by the subspace specific classifier. Each subspace's decision is then compared to the available ground truth for that sample. If a subspace's decision is correct then we increase its weight by a specified fixed amount. However, should the decision be wrong then the weight is decreased by similarly specified amount. This process is cyclically repeated over the cross-validation set until a specified iteration limit is reached.

The second method is simple least squares optimization based on the cross-validation set. For each sample we know the desired label and in the same fashion as the previous method, we can compute each subspace specific decision. Ideally, we would like a weighted linear combination that yields the correct label for each sample in the cross-validation set. Using

a least squares optimization, we can compute the weights that minimize the error between the desired labels and those computed from the weighted linear combination of the subspace specific decisions. Unlike the previous method, this one is not iterative.

Once the weights have been determined, the overall MR classifier has been trained and is ready for testing. Input test images are first subject to the MR decomposition, then feature extraction in each subspace, followed by classification in each subspace, and finished by the GDM by weighted linear combination of the subspace specific decisions. As a result of this entire process, a single decision for the original input image is made.

3.1.2 Experimental Results

To evaluate the effectiveness of their MR classification system at identification of single-tissues in H&E stained images, Chebira et al. performed multiple experiments testing different feature sets and different MR decompositions. These experiments also served to validate the authors' long-standing belief that MR is a powerful tool in many classification task. For all details of these experiments, we refer readers to [49] though we present a short summary here.

Dr. John A. Ozolek and Dr. Carlos A. Castro provided the authors with H&E stained images of teratomas as shown in Section 2.1.1. These multiple tissue images were hand segmented and labeled by Dr. Ozolek and Dr. Castro allowing the authors to extract 200×200 pixel single-tissue sub-images to compose the data set. As a result of this labeling and sub-image extraction, the data set was composed of 270 total images evenly representing 6 tissues: mesenchyme, skin, myenteric plexus, bone, necrotic tissue, and striated muscle.

The MR decompositions experimented with included multiple DWTs utilizing different Daubechies bases and their frame equivalents. The feature sets used were based on the classic Haralick texture features [31] while utilizing some nuclei specific features. A NN was used as the fundamental classifier for each subspace due to its adaptability and the authors'

past success with it. Both forms of weight learning for GDM were evaluated and compared.

To train the overall system, 40 images of the available 45 for each tissue were randomly selected. From these 40 images, cross-validation sets were extracted to not only train the subspace specific NNs but also the weights for the GDM. Once training was completed, the 5 images sequestered for testing were evaluated by the system and the correctness of the computed label determined. This division of training and testing was repeated 9 times with selection being random. The best average accuracy achieved amongst all experimental conditions was 87.7% through the use of the MR frame decomposition and the least squares optimization weight learning scheme. This was in comparison to the standard image classification scheme's accuracy of 71.7% where no MR decomposition or GDM was used.

3.1.3 Conclusions

The primary conclusion from this work was that single-tissue classification of H&E stained images of teratomas is possible at a non-trivial level of performance. Additionally, MR decomposition demonstrated its usefulness by outperforming the standard image classification model. The derived subspaces do present information that is better suited to the classification task than the original images's presentation alone. Lastly, the efficacy of texture features validated a common belief that texture is very important to the description of this family of images.

3.2 Issues to Address

Given these previous results on single-tissue classification, the next step is to address the task of multiple-tissue classification. The simple way forward from this previous work is to extend the successful single-tissue algorithm to a multiple-tissue scenario. However, the single-tissue algorithm is only designed to assign a single label to an entire image. Thus,

in order to apply it to multiple-tissue images, we would need to develop an algorithm to automate the required segmentation of a multiple-tissue image in to single-tissue regions.

It is here that our motivation for our guiding principle concerning *pixel-level classification* comes from. We argue that the very nature of tissues in H&E stained images makes such a segmentation equivalent to the desired delineation task. This is because an unsupervised segmentation of the images based on visual similarity will not yield single-tissue regions for all tissue types. Some tissues are distinguishable based on visual similarity such as color and texture, but many require higher-order descriptors such as architecture of components and spatial relationships.

Furthermore, the levels of intra-class variability and inter-class similarity with respect to visual appearance would result in significant segmentation errors. If one can achieve a supervised segmentation of the image into single-tissue regions, then the desired task is already complete. For these reasons, we propose to design a new classification system that achieves this supervised segmentation and labeling or delineation based on the classification of individual pixels and their local information.

Our other main issue concerns the feature set used to describe this family of images. Our overall goal is a system capable of addressing a wide variety of applications using H&E stained images without reinvention of the system each time. A critical component in the pursuit of this goal is the design of a core feature set that is generalizable across these applications. The design of such a feature set must be precise in its descriptions and also intuitive or any perceived levels of performance would be more difficult to explain.

While the previously used feature set based on texture did perform well, it provided no real intuition into the discriminative qualities of the image that it was using. This intuition not only provides justification for the feature design choices that we would make but also some initial understanding as to why they would or would not work in certain applications. In order to achieve this goal of a generalizable and intuitive feature set, we leverage the

expert-knowledge of the pathologist whose training, experience, and clear ability to perform the tasks of many applications lends credence to the existence of such a feature set. This approach is summarized by our *expert knowledge* guiding principle.

Chapter 4

Mathematical Framework

The first contribution of this thesis is a mathematical framework to describe through characterization of their textures, certain families of images, including the family of H&E-stained images relevant to histopathology applications. While this framework is general enough to apply to many different types of images, in the scope of this thesis it serves two main purposes: First, this framework motivates and justifies the use of many of the features we develop for our algorithm. Second, it provides a unified basis for the implementation of these features.

The fundamental components of this framework are *local histograms* and *occlusion models*. A local histogram of an image is simply the histogram of the values of the pixels in a fixed spatial neighborhood of a given pixel. With local histograms we aim to characterize various types of texture for the purpose of pixel-level classification. Occlusion models both describe and generate complex textures from simpler ones and thus extend the applicability of local histograms to more complex families of images. We will discuss these two main concepts in detail in Sections 4.2 and 4.3, respectively. In the context of our framework, the terms image and texture will be used interchangeably. Proofs for various theorems and propositions can be found in Appendix A to improve readability.

4.1   Characterization of Textures

4.1.1   Simple Textures

Pathologists use numerous visual cues when identifying tissues and pathologies (see Chapter 5 for more details). For example, some tissues primarily absorb one stain over the other. Bone's preference for eosin results in a strong pink color making the color of a tissue a good indicator for bone. Similarly, cartilage has a strong preference for hematoxylin resulting in a strong blue-purple color. However, cartilage often also contains ellipse shaped lumen regions where no stain is absorbed resulting in an overall appearance of white ellipses on top of blue-purple background (see Figure 4.1). Naturally, these statements are generalizations with specific examples of these tissues presenting variations in their exact appearances.

(a)   (b)
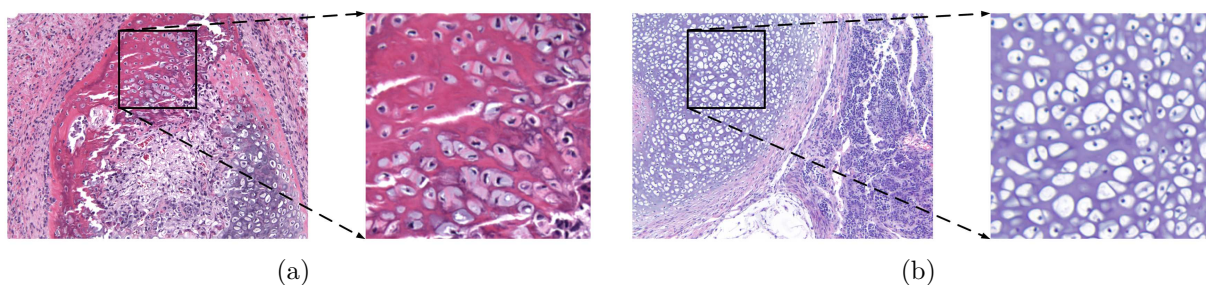
Figure 4.1: Comparison between local spatial color distribution of (a) bone and (b). On the left is the original image while on the right is an extracted sub-region of the image. Both the colors and the general shapes they present are unique enough to discriminate between the two tissue types.

Common visual qualities such as color, shape, and texture are important to the many tasks a pathologist performs. All of these cues stem from the manner in which H&E stains the various tissues and pathologies into their characteristic shades of pink, blue-purple, and white. The color of a tissue alone can be very indicative of its identity but these colors also give rise to all other visual cues that are typically considered different from color.

For example, the perceived texture of an image, e.g., smooth versus rough, is a direct result of different colors and their shades changing over the span of the image. Smooth

textures are the result of spatially constant, or slowly-changing shades of the same color, while rough textures are formed by quickly changing, or random colors. Another example is shape where it is the constancy in the color of a region, or the sharp difference between regions, which define the boundaries and span of a shape. While this is a very natural and intuitive concept, color, texture, and shape are often disconnected in their descriptions for a variety of reasons.

However, we can generally consider all of these visual cues as types of textures formed by various spatial distributions of colors or the values of the image pixels. Consider a solid pink image whose most concise description is the color pink itself. Alternatively, since pink is considered a shade of red, we could describe this same image as a completely smooth texture of that particular shade of red. Another description could be that it is formed by choosing pink for all pixels.

To restrict the family of images, we could use many methods beginning with a simple counting of how many red and blue pixels were in a given image. This description is an approximation of the image's distribution, or *global histogram* of the pixel colors. While this may be sufficient to describe certain sub-families of this already restricted family, clearly not all member images can be described in a unique manner. For example, we could group images based on the percentage of red and blue pixels present and use global histograms to describe each such group. However, we would not be able to distinguish between one image whose top and bottom halves were red and blue, respectively, and the other whose left and right halves where red and blue, respectively. Depending on the application, similar and more complex distinctions might be required. For this example, we have two extremes for possible approximations that allow us to uniquely characterize various groups within this family of images. One is the exact configuration of red and blue pixels for a given image and the other the global histogram of those same red and blue pixels. Given some set of desired groupings of the images, an optimal description lies somewhere between these extremes, providing a

balance between conciseness and the ability to distinguish between the groups. One aspect of these descriptions that allows us to travel between these extremes is the spatial nature of the description.

Depending on the image groups we want to characterize, different types of spatial localization are possible such as the location of the pixels, shape of the regions, or some combination of the two. At one extreme there is the maximum amount of spatial localization by determining each individual pixel's red or blue value, while at the other extreme of the global histogram provides no spatial localization as it does not say where in the image the red and blue pixels occur. This is very similar to the comparison between the STFT (spectrogram) and the FT where the former provides some temporal/spatial localization of frequencies while the latter provides no such localization. In fact, we can make a strong comparison between the nature and relationship of the FT and the STFT to that of a global histogram and the *local histogram* described previously. Continuing the analogy, the FT quantifies frequency over an entire signal, while a global histogram measures the frequency of occurrence for pixel values across an entire image. The STFT and local histogram introduce spatial localization by computing the same transforms as the FT and global histogram, respectively, in a spatially-dependent manner. By controlling the degree of spatial localization, we can use local histograms to effectively characterize different types of spatially dependent behaviors, i.e., textures.

Based on this motivation for the use of local histograms, we now introduce formalism that we will use throughout this chapter. We define an image as a function from a finite abelian group $\mathcal{X}$ of pixel locations into a finite abelian group $\mathcal{Y}$ of pixel values. Intuitively, this states that images are of a given spatial resolution defined by $\mathcal{X}$ and whose pixel values are drawn from some finite volume defined by $\mathcal{Y}$. As a result, our images f are members of the set $\ell(\mathcal{X}, \mathcal{Y}) := \{f : \mathcal{X} \to \mathcal{Y}\}$. For example, the 1200×1600, 8-bit RGB images from Figure 1.3 have $\mathcal{X} = \mathbb{Z}_{1200} \times \mathbb{Z}_{1600}$ and $\mathcal{Y} = \mathbb{Z}_{256}^3$, where $\mathbb{Z}_N$ denotes the cyclic group of

integers modulo N.

The local histogram of an image f is defined in terms of a *weighting function* or *neighborhood w* that dictates the local histogram's spatial localization. More specifically, $w \in \ell(\mathcal{X}, \mathbb{R})$, is non-negative, and sums to one. Thus, the *local histogram transform* of the image f with respect to a weighting function w, is the function $\mathrm{LH}_w f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,

$$(\mathrm{LH}_w f)(x, y) := \sum_{x' \in \mathcal{X}} w(x') \delta_y(f(x + x')). \tag{4.1}$$

where

$$\delta_y(f(x)) = \begin{cases} 1, & f(x) = y, \\ 0, & f(x) \neq y, \end{cases} \tag{4.2}$$

For any fixed $x \in \mathcal{X}$ (pixel location) and $y \in \mathcal{Y}$ (pixel value), $(\mathrm{LH}_w f)(x, y)$ counts the number of pixels for which f has the value y in the neighborhood of x defined by w.

Comparing bone to cartilage (see Figure 4.1), our intuition is that the color local histogram of bone should have peaks in the pink region while the color local histogram of cartilage will have peaks in the blue and white regions, based on which, it is relatively easy to distinguish between the two tissues by comparing the number and location of their peaks.

4.1.2 Complex Textures

While local histograms can characterize certain types of textures well, there will be complex textures that will be difficult to understand through local histograms alone, such as complex textures that still possess structure. By structure we refer to textures that are not the result of purely random assignment of pixel values; this is relevant to our applications since structure is present in the form of various biological components.

Our question now is whether we can use simplex textures and their corresponding local histograms to characterize complex ones. We propose to use *occlusion models* for that task.

In Section 4.3, we study occlusion models in detail; here, we provide some intuition.

Let us revisit our previous example of the family of images created by choosing either red or blue for each pixel in an image. Suppose that selection of the pixel color was done according to some independent and identically distributed (i.i.d.) distribution for each pixel, specifically, a Bernoulli distribution with probability $p \in [0,1]$ that the color red is chosen and probability $(1-p)$ that blue is chosen. Now let us compute its local histograms using some fixed w over their red and blue values. We expect that, on average, for any image, the values of $(\mathrm{LH}_w f)(x, \cdot)$ for any $x \in \mathcal{X}$ reflect this probability distribution in choosing red or blue. In other words, for y representing red and blue, $(\mathrm{LH}_w f)(x, y)$ should on average equal p and $(1-p)$, respectively.

Now let us change our image formation model to one beginning with two images, a solid red one f_0 and a solid blue one f_1. Our final image f is formed by *occluding* the blue image with the red image at certain locations; specifically, those locations where a random Bernoulli trial picks red with probability p. This is equivalent to the previous image formation model but we now form our image as a function of other images. More generally, the *occlusion* of a set of N images $\{f_n\}_{n=0}^{N-1}$ in $\ell(\mathcal{X}, \mathcal{Y})$ with respect to a given *label function* $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)$ is

$$\left(\mathrm{occ}_\varphi \{f_n\}_{n=0}^{N-1}\right)(x) := f_{\varphi(x)}(x). \tag{4.3}$$

That is, at any pixel location x, the label $\varphi(x)$ determines which of the available pixel values $\{f_n(x)\}_{n=0}^{N-1}$ will appear in the composite image $\mathrm{occ}_\varphi \{f_n\}_{n=0}^{N-1}$ at that location.

Note that φ is deterministic and results in a specific occlusion, while the occlusion model Φ is its random variable version. To be precise, fix a set of source images $\{f_n\}_{n=0}^{N-1}$ and consider the set $\{\mathrm{occ}_\varphi \{f_n\}_{n=0}^{N-1}\}_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)}$ of all possible composite images (4.3) obtained by letting φ be any one of the $N^{|\mathcal{X}|}$ elements of $\ell(\mathcal{X}, \mathbb{Z}_N)$, where $|\mathcal{X}|$ denotes the cardinality of $\mathcal{X}$.

In our example, Φ is a sequence of i.i.d. Bernoulli trials and, as a result, each φ has

an associated probability of occurrence $P_\Phi(\varphi) \in [0,1]$. Thus, our occlusion model Φ has an associated probability density function $P_\Phi : \ell(\mathcal{X}, \mathbb{Z}_N) \to [0,1]$ such that $\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) = 1$. It is the role of the associated probability P_Φ to assign probabilities to those occlusions φ that will result in images relevant to the application at hand. In our example, of all possible occlusions φ that can be generated by Φ, only a subset will generate images that do not look like mere red-blue noise. By assigning low probabilities to these noise generating occlusions, the occlusion model will emphasize desirable occlusions while de-emphasizing undesirable ones.

Given this method of forming our image, it is clear that the source images f_n's and a fixed φ determine the composite image. In our simple red and blue example, the color local histograms of f_0 and f_1 each contain only one peak in the red and blue regions, respectively. Based on this occlusion model we would expect that the color local histogram of the composite image $\mathrm{LH}_w \mathrm{occ}_\varphi \{f_0, f_1\}$ should be some combination of the local histograms of the source images. Specifically, given the probability p of choosing red, $\mathrm{LH}_w \mathrm{occ}_\varphi \{f_0, f_1\}$ should be $p\mathrm{LH}_w f_0 + (1-p)\mathrm{LH}_w f_1$, on average. Our intuition here is that the local histogram of the composite image is some convex combination of the local histograms of the source images.

This toy example illustrates the basic concept we wish to generalize to more realistic and relevant families of images and textures. Given a probabilistic occlusion model with which to form a composite image from a set of source images, what is the relationship between the local histograms of the source images and that of the composite image on average? This relationship will be rigorously analyzed so as to determine under what conditions both strong and weak statements can be made. The strength of such statements in various applications will support expectations about the usefulness of local histograms in such applications.

In our example, occlusion models can generate images, and, in fact, any image can be generated through the action of occlusion. If one considers our set of source images $\{f_n(x)\}_{n=0}^{N-1}$ as being solid-color images of every possible color, any given image is merely

a particular occlusion drawn from all possible occlusions. Of course, the set of all possible occlusions will lead to images that are completely random in appearance and do not represent any real image. Realistic images will correspond to those occlusions where spatial correlations and an occlusion hierarchy exist at various degrees.

Spatial correlations refers to the situations where a group of pixel locations (those spatially connected to each other) will together take on the pixel values from one particular source image f_n. Such groups of pixels are in fact the shapes, layers, and objects we perceive in images due to their constant or structured appearance. The spatial distribution, i.e., shapes of such groups is important to creating composite images that correspond well to actual images. For example, consider a source image representing wood's appearance. The objects made using this wood are often distinguished through their shape and thus in an occlusion model such considerations must be made to generate realistic images.

Occlusion hierarchy refers to those situations in which certain source images should only occlude certain other source images for contextual reasons. A simple example is that of the human face and its component parts. If we were to create an image of a human face, our source images could be those of the basic parts of a face. These include the skin-colored face shape, the mouth, the nose, ears, eyes, and more. Naturally, we would occlude the image of the colored face shape with the other parts since that is where they belong. However, when placing the mouth and the nose, we would not want them to occlude each other. If, additionally, we separated the eyes into just the outline of the eyes and the iris/pupil region, the iris should only occlude the outline of the eye and no other region. In other words, given a set of source images that reflect the context of the images we wish to generate, there is likely an order or hierarchy that helps ensure that the occlusion model based generation of composite images is realistic within that context.

In this chapter, we present a rigorous explanation and discussion of the relationship between local histograms and occlusion models. To that end, we will provide theoretical

justification for both strong and weak statements about the applicability of local histograms of images that can be characterized by occlusion models. Furthermore, our discussion about occlusion models will present operations which can be used to generate complex textures and images from simpler ones while allowing us to relate them to local histograms. Finally, we present some considerations about computational requirements and efficiency in the context of our histopathology applications.

4.2 Local Histograms

Local histograms become increasingly time-consuming to compute as the image's spatial resolution $\mathcal{X}$ and its value resolution $\mathcal{Y}$ increase. A simple implementation of (4.1) would involve sequentially addressing each pixel location and collecting their neighboring pixel values (as dictated by the specified w). Such a computation would require $\mathcal{O}(|\mathcal{X}|^2|\mathcal{Y}|)$ operations: $\mathcal{O}(|\mathcal{X}|)$ operations for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Clearly this computation becomes intractable as the sizes of $\mathcal{X}$ and $\mathcal{Y}$ become large.

In order to make such a computation feasible, we propose the method detailed in Theorem 1 below which states that (4.1) can be computed using $|\mathcal{Y}|$ convolutions over $\mathcal{X}$, which only requires $\mathcal{O}(|\mathcal{X}||\mathcal{Y}|\log|\mathcal{X}|)$ operations if discrete Fourier transforms (DFT) are used. In particular, we filter the *characteristic function* of the graph of f, namely $1_f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,

$$1_f(x,y) := 1_{f^{-1}\{y\}}(x) = \delta_y(f(x)) = \begin{cases} 1, & f(x) = y, \\ 0, & f(x) \neq y, \end{cases} \tag{4.4}$$

with the *reversal* of $w \in \ell(\mathcal{X}, \mathbb{R})$, namely $\tilde{w}(x) := w(-x)$. Theorem 1(b) also states that (4.1) can be computed as a single convolution over $\mathcal{X} \times \mathcal{Y}$ as opposed to a series of convolutions. The tensor product of $w \in \ell(\mathcal{X}, \mathbb{R})$ with the $\omega \in \ell(\mathcal{Y}, \mathbb{R})$ is defined as $w \otimes \omega \in \ell(\mathcal{X} \times \mathcal{Y}, \mathbb{R})$, $(w \otimes \omega)(x,y) := w(x)\omega(y)$.

Theorem 1. *For any $w \in \ell(\mathcal{X}, \mathbb{R})$, $\omega \in \ell(\mathcal{Y}, \mathbb{R})$, $f \in \ell(\mathcal{X}, \mathcal{Y})$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$:*

(a) *Local histograms (4.1) can be evaluated as a system of $|\mathcal{Y}|$ convolutions over $\mathcal{X}$:*

$$(\mathrm{LH}_w f)(x, y) = (\tilde{w} * 1_{f^{-1}\{y\}})(x).$$

(b) *Alternatively, (4.1) may be computed as a single convolution over $\mathcal{X} \times \mathcal{Y}$: $(\delta_0 \otimes \omega) * \mathrm{LH}_w f = (\tilde{w} \otimes \omega) * 1_f$. In particular, taking $\omega = \delta_0$ gives $\mathrm{LH}_w f = (\tilde{w} \otimes \delta_0) * 1_f$.*

Figure 4.2 shows the computation of a simple image's local histograms using Theorem 1(a). The methods of computation in Theorem 1 are based on basic filtering concepts, and thus, we might expect that some related signal processing concepts may also apply to local histograms. Specifically, we look at the effects of spatial translations, constant value shifts, and value quantizations of the input image on on local histograms. These properties are summarized below.

Proposition 2. *For any $w \in \ell(\mathcal{X}, \mathcal{R})$ and $f \in \ell(\mathcal{X}, \mathcal{Y})$:*

(a) *The levels of a local histogram transform sum to one: for any $x \in \mathcal{X}$,*

$$\textstyle\sum_{y \in \mathcal{Y}} (\mathrm{LH}_w f)(x, y) = 1.$$

(b) *Local histograms commute with spatial translation: for any $x \in \mathcal{X}$, $\mathrm{LH}_w \mathrm{T}^x = \mathrm{T}^{(x,0)} \mathrm{LH}_w$.*

(c) *Adding constants to images shifts their local histograms along $\mathcal{Y}$: for any $y \in \mathcal{Y}$,*

$$\mathrm{LH}_w(f + y) = \mathrm{T}^{(0,y)} \mathrm{LH}_w f.$$

(d) *Quantizing an image will bin its local histograms: for any $q \in \ell(\mathcal{Y}, \mathcal{Y}')$,*

$$[\mathrm{LH}_w(q \circ f)](x, y') = \sum_{\substack{y \in \mathcal{Y} \\ q(y) = y'}} (\mathrm{LH}_w f)(x, y).$$

With the ability to compute local histograms in reasonably efficient manner, we now proceed to discuss their relationship to simple probabilistic occlusion models and the textures that can be generated by them.
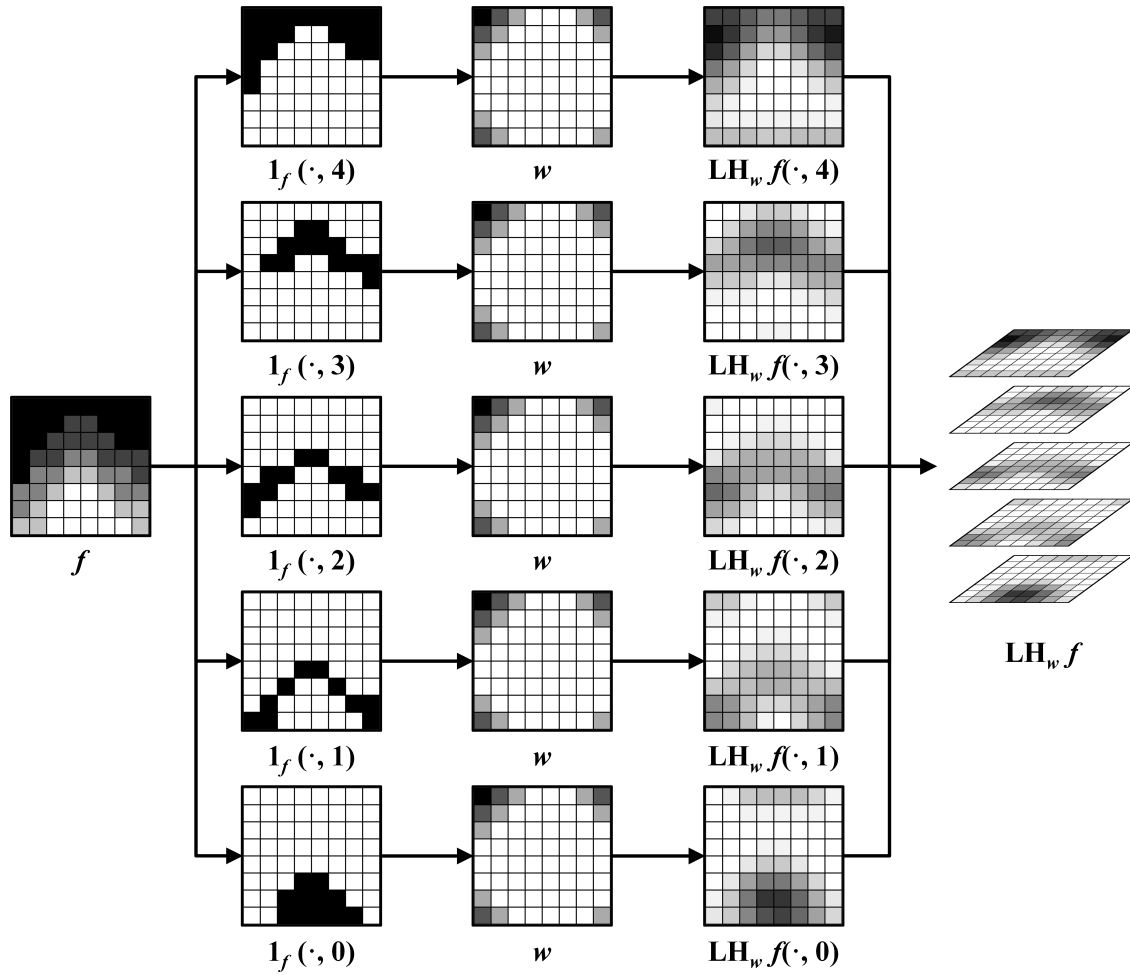
Figure 4.2: Example of a simple image's local histogram computation using Theorem 1(a). For clarity purposes, larger values are represented by darker shades. The input image f (far-left/first column) is an 8 pixel image whose grayscale values range from 0 to 4 or equivalently, $f \in \ell(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X} = \mathbb{Z}_8 \times \mathbb{Z}_8$ and $\mathcal{Y} = \mathbb{Z}_5$. The characteristic function $1_f(x, y)$ as defined by (4.4) is a $8 \times 8 \times 5$ binary matrix whose cross-sections are $1_f(\cdot, y)$ (second column). For a given value of $y \in \mathcal{Y}$, $1_f(\cdot, y)$ assumes a value of 1 at the locations x for which $f(x) = y$ which are represented by black pixels in each cross-section. Following Theorem 1(a), each of these cross-sections is then filtered with the specified real-valued weighting function w (third column) to yield the corresponding cross-section of the local histogram $(\mathrm{LH}_w f)(\cdot, y)$ (fourth column). Collectively these cross-sections form the $8 \times 8 \times 5$ matrix representing f's local histogram $(\mathrm{LH}_w f)(x, y)$ (far-right/last column). In this simple example, the computation is done using DFTs and thus results in circular convolution as opposed to linear convolution the effects of which can be seen in values of the local histogram cross-sections near their boundaries. However, such issues can be mitigated using many available techniques including zero-padding and overlap-add/save methods. In this example, $w = \frac{3}{19}\delta_{0,0} + \frac{2}{19}(\delta_{-1,0} + \delta_{1,0} + \delta_{0,-1} + \delta_{0,1}) + \frac{1}{19}(\delta_{-2,0} + \delta_{2,0} + \delta_{0,-2} + \delta_{0,2} + \delta_{-1,-1} + \delta_{1,1} + \delta_{-1,1} + \delta_{1,-1})$.

4.3 Occlusion Models

In this section we will discuss an image or texture formation model that is capable of creating families of images well suited to analysis by local histograms. We start with random textures and incrementally introduce both spatial and contextual structure. Specifically, complex textures will be generated as functions of simpler ones. As the complexity of the textures increase, we will rigorously analyze their relationship to local histograms to the point that they reflect real families of images, specifically those of the textures found in H&E-stained histopathology images.

4.3.1 Random Textures

We wish to confirm our intuition that if a composite texture is the result of some sufficiently random occlusion model, then its local histograms are on average equivalent to some convex combination of the local histograms of the simpler source textures. As stated earlier, given such a random occlusion model, it would be unrealistic to expect a precise relationship and we expect there to be some uncertainty in this convex combination. However, in later sections we will show under what conditions this uncertainty can be reduced and even eliminated to provide a clear relationship between composite and source images.

To begin, consider a fixed set of N source images $\{f_n\}_{n=0}^{N-1}$ in addition to a random occlusion model Φ. As described earlier, Φ is the random variable equivalent of the deterministic label function $\varphi : \mathcal{X} \to \mathbb{Z}_N$; each particular label function φ represented by Φ has an associated probability of occurrence. Thus, Φ is characterized by a proper probability $P_\Phi : \ell(\mathcal{X}, \mathbb{Z}_N) \to [0, 1]$ where $\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_\Phi(\varphi) = 1$.

Consider the expected value, with respect to P_Φ, of the characteristic function 1_φ derived

from setting $f = \varphi$ in (4.4):

$$\overline{1}_{\Phi}(x, n) := \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} P_{\Phi}(\varphi) 1_{\varphi}(x, n) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \varphi(x)=n}} P_{\Phi}(\varphi). \tag{4.5}$$

Essentially, $\overline{1}_{\Phi}(x, n)$ is the probability that any label function φ randomly generated/selected by the occlusion model Φ will assign label n to pixel location x. Using this characterization of the random occlusion Φ, we now present one of the main results of this mathematical framework:

Theorem 3. *For any sequence of images $\{f_n\}_{n=0}^{N-1} \in \ell(\mathcal{X}, \mathcal{Y})$, weighting function w and any N-image occlusion model Φ, the expected value of the local histogram (4.1) of the composite image (4.3) with respect to w is:*

$$E_{\Phi}(\mathrm{LH}_w \mathrm{occ}_{\Phi} \{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \overline{1}_{\Phi}(x, n)(\mathrm{LH}_w f_n)(x, y) + \varepsilon, \tag{4.6}$$

where the error term ε is bounded by $|\varepsilon| \leq \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x')|\overline{1}_{\Phi}(x + x', n) - \overline{1}_{\Phi}(x, n)|$. Moreover,

$$\sum_{n=0}^{N-1} \overline{1}_{\Phi}(x, n) = 1, \tag{4.7}$$

and so (4.6) states that, on average, the local histograms of the composite image $\mathrm{occ}_{\varphi} \{f_n\}_{n=0}^{N-1}$ can be approximated by convex combinations of local histograms of each individual image f_n.

Figure 4.3 presents a simple example of computing the left-hand side of (4.6) explicitly.

Theorem 3 implies that the error term ε in (4.6) will be small when the probability $\overline{1}_{\Phi}(x, n)$ of assigning label n to x changes little as x varies over regions smaller than the the support of w. As ε approaches 0, (4.6) provides a increasingly clean relationship between the local histograms of a composite image and those of its source images. Given a fixed set
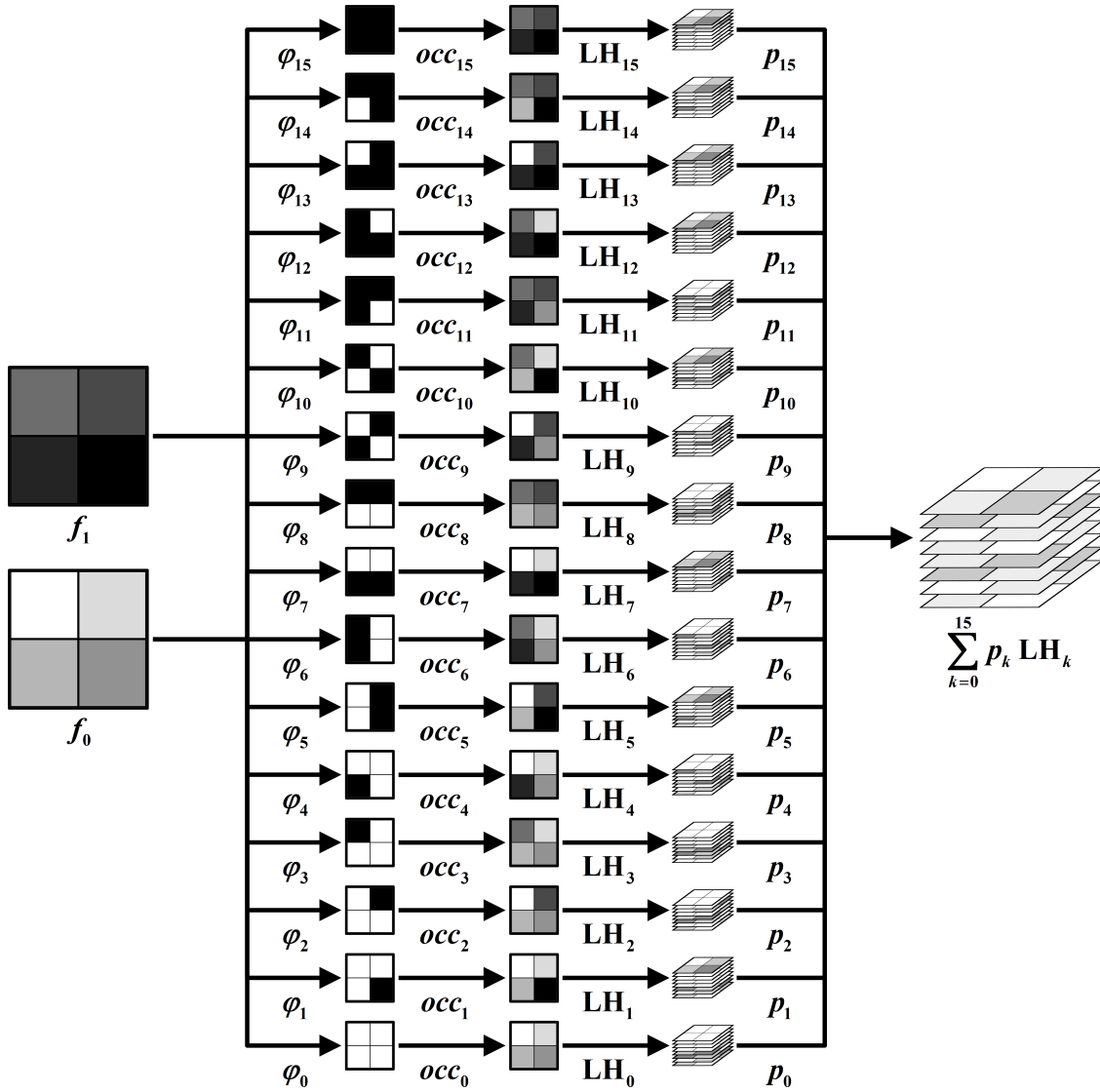
Figure 4.3: A simple example of how to compute the left-hand side of (4.6) explicitly as a probability-weighted sum. For clarity purposes, larger values are represented by darker shades. We begin with two 2×2, 3-bit source images (far-left/first column), specifically $\{f_n\}_{n=0}^{N-1}$ in $\ell(\mathcal{X}, \mathcal{Y})$ where $N = 2$, $\mathcal{X} = \mathbb{Z}_2 \times \mathbb{Z}_2$ and $\mathcal{Y} = \mathbb{Z}_8$. The pixel values of the f_n's in this example are all distinct such that f_0 takes on values $\{0, 1, 2, 3\}$ while f_1 has values $\{4, 5, 6, 7\}$. There are $N^{|\mathcal{X}|} = 2^{2^2} = 16$ distinct ways or label functions $\varphi : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2$ with which to occlude one image with another (second column). Following (4.3), we take values from f_0 in places where φ is white (representing a 0 label) and values from f_1 where φ is black (representing a 1 label) with each distinct φ yielding a composite image $\mathrm{occ}_\varphi\{f_0, f_1\}$ (third column). Each of these composites has a local histogram transform (4.1) of size $2 \times 2 \times 8$ (fourth column). Based on the probability $\mathrm{P}_\Phi(\varphi)$ of any particular label function φ occurring, the expected value of the local histogram transform of the composite image can be computed (far-right/last column).

of source images $\{f_n\}_{n=0}^{N-1}$ and weighting function w, the value of ε is entirely dependent on

$\bar{\mathsf{I}}_\Phi(x, n)$ and therefore on the occlusion model Φ. Due to this dependence, the conditions un-

der which ε vanishes entirely are conditions on Φ; specifically, we consider a special condition

where Φ is *flat*.

Flatness means that on average, the probability that Φ chooses label n at a given pixel

location x is equal to the probability of choosing n at any other x'. Formally, Φ is *flat* if

there exists scalars $\{\lambda_n\}_{n=0}^{N-1}$ such that:

$$\sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N) \\ \varphi(x) = n}} \mathrm{P}_\Phi(\varphi) = \lambda_n, \quad \forall x \in \mathcal{X}. \tag{4.8}$$

Equivalently, Φ is flat if the marginal distributions obtained by fixing any given $x \in \mathcal{X}$

are identical. In fact, for any fixed $x \in \mathcal{X}$, the summation of (4.8) over all n results in

$\sum_{n=1}^{N} \lambda_n = 1$. Indeed, at any given pixel location x, the value λ_n is the probability that any

random label function φ from the occlusion model Φ will have label n at that x.

Revisiting our red-blue pixel example in Section 4.1, the values of any given $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$

are determined by $|\mathcal{X}|$ independent Bernoulli trials. As a result, the probability of any

particular φ's configuration of labels is $\mathrm{P}_\Phi(\varphi) = p^{|\varphi^{-1}\{1\}|}(1-p)^{|\mathcal{X}|-|\varphi^{-1}\{1\}|}$. Substituting this

probability into (4.8), the binomial theorem implies that this occlusion model is flat with

$\lambda_0 = p$ and $\lambda_1 = 1 - p$. An important distinction is that while this model is flat, it does

not imply that the contributions from each of the source images to the composite image are

equal. In our example, if $p > 0.5$ then the resulting random composite image $\mathrm{occ}_\Phi\{f_0, f_1\}$

from our flat occlusion model Φ will be more red than blue. Rather, it indicates that the

likelihood of a given pixel being red at any given location is equal to the likelihood at any

other location under the occlusion model.

Given these ideas, we present another claim of this mathematical framework: that on

average, composite images generated by flat occlusion models will be convex combinations

of the source images' local histograms.s

Theorem 4. *If Φ is flat (4.8), then the expected value of the local histograms (4.1) of a composite image (4.3) is a convex combination of the local histograms of each individual image:*

$$\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi)(\mathrm{LH}_w \mathrm{occ}_\varphi \{f_n\}_{n=0}^{N-1})(x,y) = \sum_{n=0}^{N-1} \lambda_n (\mathrm{LH}_w f_n)(x,y). \tag{4.9}$$

That is, when Φ is flat, (4.6) simplifies to (4.9). The explicit computation presented in Figure 4.3 can be simplified to the one shown in Figure 4.4. Flatness is an important theoretical assumption for analysis of textures by local histograms; we now must show in what situations flatness is a realistic assumption. Specifically, we discuss which types of images and textures can be characterized by flat occlusion models.
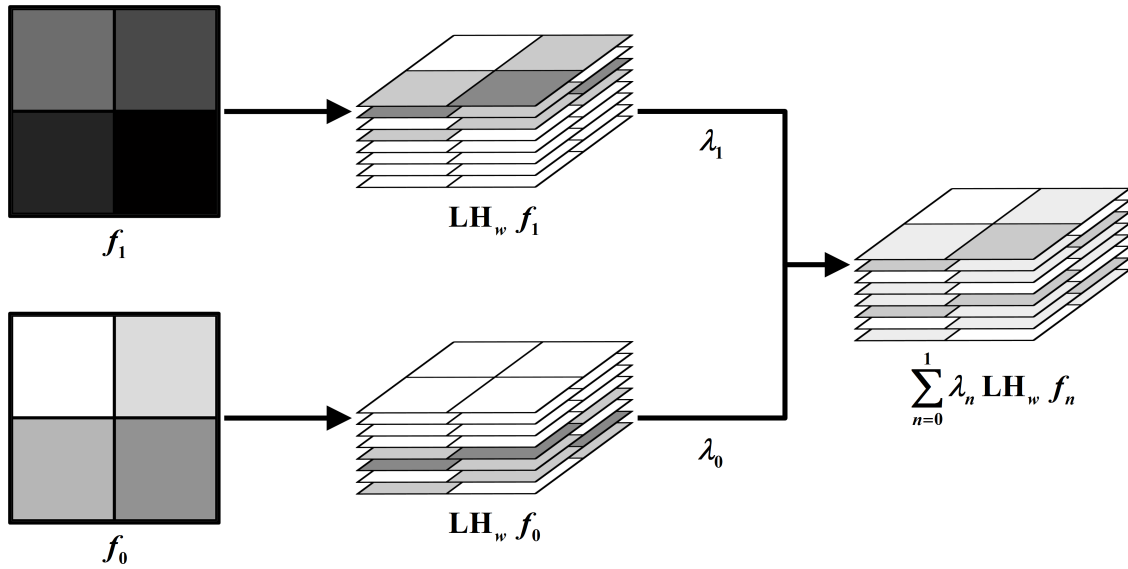
Figure 4.4: An extension of the example presented in Figure 4.3 where the occlusion model is flat. Instead of the explicit computation of the expected value of the composite image's local histogram (4.6), the flatness of the occlusion model allows us to simplify the computation. From each of the source images f_0 and f_1 (far-left/first column) their local histograms $LH_w f_0$ and $\mathrm{LH}_w f_1$ (second column) can be computed respectively. Based on (4.9), the expected value of the composite image's local histogram can be computed from the source images' local histograms.

4.3.2 Flat Occlusion Models

In the previous section we analyzed the relationship between the local histograms of composite images and random occlusion models that generated them. This relationship began relatively weak in Theorem 3 but with the condition of *flatness* (4.8) imposed, became much stronger in Theorem 4. Flat occlusion models are those whose average characteristic function $\overline{1}_\Phi(x, n)$, as defined by (4.5), is constant with respect to pixel location x, but can still vary with label value n.

We want to show that flatness is a reasonable assumption, or equivalently that certain families of images and textures can indeed be modeled by flat occlusion models. As our definition of occlusion (4.3) is one which generates images, we will put forth a variety of methods with which to construct flat models. We begin with a property of occlusion models that implies flatness or more specifically, that if an occlusion model Φ is *translation-invariant* then it is flat. Formally, given the translation operator $T^x : \mathcal{X} \to \mathcal{X}$, $T^x\varphi(x') := \varphi(x' - x)$, a translation-invariant occlusion model obeys:

$$P_\Phi(T^x\varphi) = P_\Phi(\varphi), \quad \forall \varphi \in \ell(\mathcal{X}, \mathbb{Z}_N), x \in \mathcal{X}. \tag{4.10}$$

Theorem 5. *If Φ is translation-invariant* (4.10)*, then Φ is flat* (4.8)*.*

Theorem 5 demonstrates that flatness is not an unrealistic assumption. If we generalize our red-blue example of Section 4.1 to one that merely picks a number from $\mathbb{Z}_N$ based on some random method, φ is generated by conducting $|\mathcal{X}|$ independent trials. Under this occlusion model Φ, the probability for any particular occlusion $P_\Phi(\varphi)$ is completely determined by the number of trials for which φ achieves each given value n. As another example, consider a fixed φ_0 and all of its translates, to each of which we can assign identical non-zero probability $\frac{1}{|\mathcal{X}|}$ and probability 0 to all others. If we apply this occlusion model to a set of source images $\{f_n\}_{n=0}^{N-1}$ that are constant then the resulting composite images (4.3) are all translations of

a single image.

In general, given a fixed resolution $|\mathcal{X}|$ and a fixed number of source images N, all possible $N^{|\mathcal{X}|}$ occlusions of $\ell(\mathcal{X}, \mathbb{Z}_N)$ can be grouped into translation-invariant groups (equivalence classes). For all the members of each of these equivalence classes one can assign a fixed probability, provided they all sum to one, and thus make the occlusion model flat according to Theorem 5. Figure 4.3 shows an example where $N = 2$ and $\mathcal{X} = \mathbb{Z}_2 \times \mathbb{Z}_2$ in which we may partition the 16 possible φ's into 7 such equivalence classes. To make this occlusion model flat, one would pick any probabilities $\{p_k\}_{k=0}^{15}$ such that $p_1 = p_2 = p_3 = p_4$, $p_5 = p_6$, $p_7 = p_8$, $p_9 = p_{10}$, $p_{11} = p_{12} = p_{13} = p_{14}$.

Translation invariance provides our first method for producing flat occlusion models Φ from the general occlusion model for a given resolution $\mathcal{X}$ and source images $\{f_n\}_{n=0}^{N-1}$. Translation-invariant occlusion models may be sufficient for certain families of images/texture such as some of our histopathology-related tissues, but for others they may not be sufficient or even appropriate. To widen the applicability of occlusion models, we now discuss methods of combining known occlusion models to produce more complex and realistic ones.

Expansion

Our next method is based on an observation of the nature of basic components in H&E stained images of tissues. As can be seen in many of the previous figures, seemingly randomly distributed blobs of varying colors make up much of a tissue's appearance. Specifically, these blobs correspond to distinct and basic biological structures such as cells and nuclei. At any given stage of development or maturity, these structures in a given tissue are the result of many cycles of growth and reproduction. As would be expected, the particular role of a tissue ensures that the distribution of these structures is near uniform, both spatially and in terms of appearance. In the case of cells, cells prefer to grow and reproduce in empty space, i.e., they do not overlap. Also, to support a tissue's main function collectively they assume

relatively constant size, shape, and appearance.

Given our histopathology application, we want to apply our occlusion models in a way that emulates the textures of tissues. If these occlusion models were flat, Theorem 4 would justify our use of local histograms in describing tissues. Such local histograms would represent a decomposition or demixing of a tissue into contributions from each of these basic components such as cells and nuclei, the intuition being that individual tissues present sufficiently distinct contributions from each of these components and thus allow us to distinguish between them based on the local histograms.

Continuing with our biological motivation, we propose a method based on the growth of these blobs beginning with a randomly chosen location. Given a way to randomly generate or *seed* a set of almost uniformly-distributed points, we can *expand* each of these points into a blob. One obvious method to create these uniformly-distributed points is to conduct a Bernoulli trial at each location.

Let $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$ indicate a set of randomly generated points. For each of the points $x \in \mathcal{X}$ for which $\varphi(x) = 1$, we will replace it with a blob whose shape is indicated by some $\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)$. The resulting texture will be the union of all such blobs. Formally, given any $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)$ and $\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}$, we define the *expansion* of φ by $\{\psi_x\}_{x \in \mathcal{X}}$ to be $\varphi \star \{\psi_x\}_{x \in \mathcal{X}} \in \ell(\mathcal{X}, \mathbb{Z}_2)$,

$$(\varphi \star \{\psi_{x'}\}_{x' \in \mathcal{X}})(x) := \begin{cases} 1, & x = x' + x'', \varphi(x') = 1, \psi_{x'}(x'') = 1, \\ 0, & \text{else.} \end{cases} \tag{4.11}$$

An important distinction is that expansion (4.11) is a deterministic operation on label functions in $\ell(\mathcal{X}, \mathbb{Z}_2)$. This is in contrast to an occlusion model which is random variable Φ defined by a probability density function P_Φ over $\ell(\mathcal{X}, \mathbb{Z}_2)$. Examples of this expansion operation are shown in Figure 4.5.

In spite of this difference, application of the expansion operation (4.11) on the random

(a) φ (b) Examples of ψ_x (c) $\varphi \star \{\psi_x\}_{x \in \mathcal{X}}$

(d) φ' (e) Examples of ψ'_x (f) $\varphi' \star \{\psi'_x\}_{x \in \mathcal{X}}$
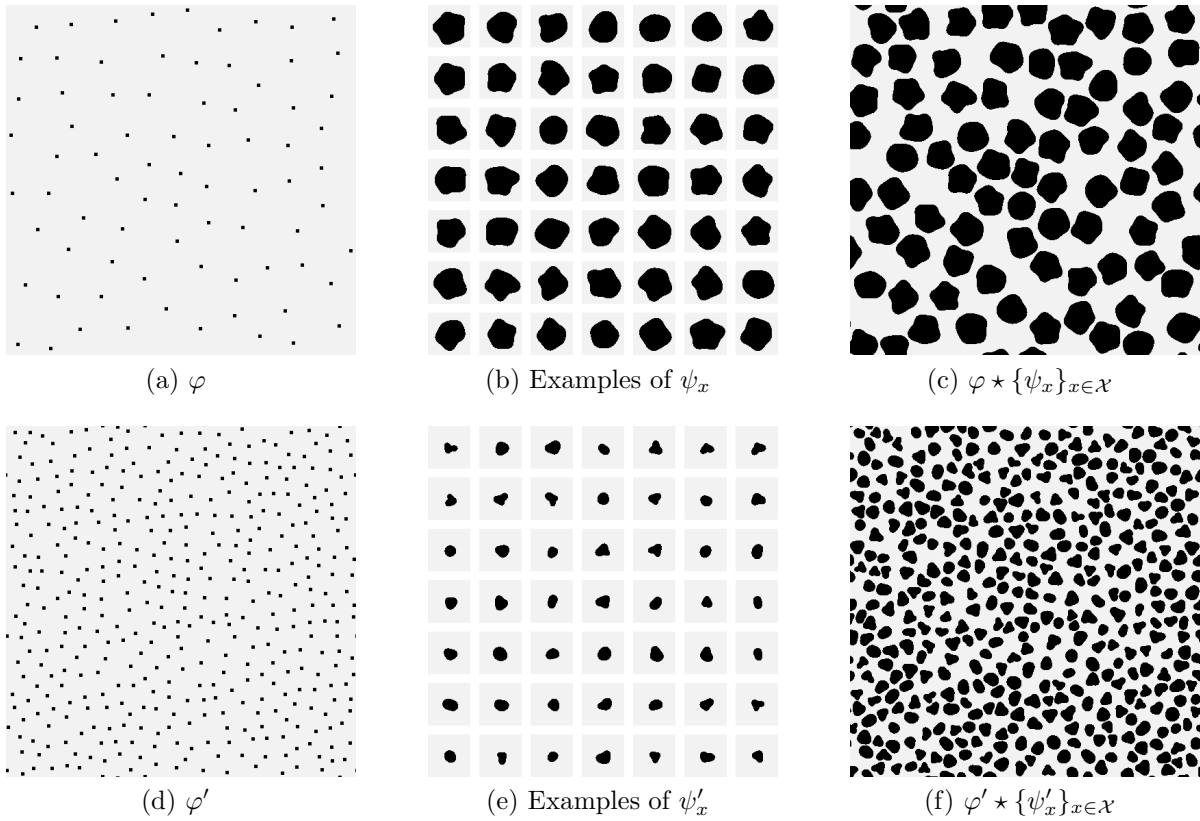
Figure 4.5: Examples of the expansion operation (4.11). For the sake of clarity, black indicates a value of 1 while the lighter shade indicates a 0 value. A function $\varphi : \mathcal{X} \to \{0, 1\}$ is presented in (a), and can be chosen, for example, via a sequence of $|\mathcal{X}|$ independent coin flips. Meanwhile, for each $x \in \mathcal{X}$, we pick a corresponding function $\psi_x : \mathcal{X} \to \{0, 1\}$. Cropped versions of a few examples of such ψ_x's are given in (b). The expansion $\varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ of φ by $\{\psi_x\}_{x \in \mathcal{X}}$ is given in (c). Essentially, each point x for which $\varphi(x) = 1$ is replaced with the corresponding blob ψ_x, with the origin of the ψ_x coordinates being translated to x. In the second row, (f) shows the expansion of a second set of points φ' (d) by a second set of blobs $\{\psi'_x\}_{x \in \mathcal{X}}$ (e). These examples notwithstanding, note that (4.11) does not require these blobs to be disjoint (4.13). We could have, for instance, produced a texture by expanding the points in (d) by the blobs in (b). Nevertheless, stronger conclusions can be made if such disjointness is enforced; see Theorem 6.

variable equivalents Φ and Ψ of the label functions φ and $\{\psi_x\}_{x\in\mathcal{X}}$ does produce a similar result. Specifically, given two occlusion models Φ and Ψ from $\mathcal{X}$ into $\mathbb{Z}_2$, we define the expansion of Φ by Ψ to be the occlusion model $\Phi \star \Psi$ whose probability density function is $P_{\Phi\star\Psi} : \ell(\mathcal{X}, \mathbb{Z}_2) \to [0, 1]$,

$$P_{\Phi\star\Psi}(\sigma) := \sum_{\substack{\varphi\in\ell(\mathcal{X},\mathbb{Z}_2) \\ \{\psi_x\}_{x\in\mathcal{X}}\in[\ell(\mathcal{X},\mathbb{Z}_2)]^{\mathcal{X}} \\ \varphi\star\{\psi_x\}_{x\in\mathcal{X}}=\sigma}} P_\Phi(\varphi) \prod_{x\in\mathcal{X}} P_\Psi(\psi_x). \tag{4.12}$$

We will prove that this is a true probability in Theorem 6. Note that the probability that $\Phi\star\Psi$ will produce a particular label function σ depends on the ways in which σ can be written as $\varphi\star\{\psi_x\}_{x\in\mathcal{X}}$ and also on the probabilities that Φ and Ψ will produce those involved φ's and ψ_x's, respectively. In other words, many different configurations of seed points and blobs can result in the same label function though some may be more or less likely due to the probabilities associated with them. It is this ambiguity in the exact parameters that lead to the construction of any given label function σ through the expansion of Φ by Ψ that we will discuss next.

We will show that if Φ is translation-invariant (4.10), then $\Phi \star \Psi$ is also translation-invariant which in turn implies that it is flat by Theorem 5. In the context of our tissue textures, occlusion models formed by expansion will be flat if the distribution that produces the seeds of the blobs is translation-invariant. In fact, to achieve flatness of $\Phi \star \Psi$ we can lessen our requirement on Φ from being translation-invariant to only requiring that Φ is flat provided that Φ and Ψ are *effectively disjoint*:

$$\text{If } P_\Phi(\varphi) > 0 \text{ and } P_\Psi(\psi_x) > 0 \text{ for all } x \in \mathcal{X}, \text{ then } \varphi \star \{\psi_x\}_{x\in\mathcal{X}} = \sum_{\substack{x\in\mathcal{X} \\ \varphi(x)=1}} \mathrm{T}^x \psi_x. \tag{4.13}$$

Effective disjointness is the requirement that there is only one way, with nontrivial probability, in which the x in (4.11) can be written as $x = x' + x''$ where both $\varphi(x') = 1$ and

$\psi_{x'}(x'') = 1$. The previously discussed ambiguity of label functions created by expansion is significantly reduced by (4.13) since the number of ways to construct a given label function is much smaller.

Theorem 6. *If Φ and Ψ are occlusion models from $\mathcal{X}$ into $\mathbb{Z}_2$, then their expansion $\Phi \star \Psi$, with probability density function (4.12), is as well. Moreover, if Φ is translation-invariant (4.10), then $\Phi \star \Psi$ is translation-invariant. Furthermore, if Φ and Ψ are effectively disjoint (4.13) and either Φ or Ψ is flat (4.8), then $\Phi \star \Psi$ is flat.*

We have shown a more complex operation of *expanding* random seeds into blobs that mimics the development of biological structures. Furthermore, the simple notion that two distinct structures cannot occupy the same space allows us to *effectively disjoin* the seeds and blobs in certain cases. This property in addition to *translation invariance* and *flatness* allows us to use this operation to create flat occlusion models that are complex enough to model relevant textures. However, expansion may still be insufficient to characterize even more complex textures of concern and thus we propose another operation to further increase the applicability of our occlusion models while retaining their link to local histograms.

Overlay

The expansion operator (4.11) allows us to combine two binary occlusion models, Φ and Ψ, into a new occlusion model $\Phi \star \Psi$ that generates random label functions according to (4.12). Furthermore, Theorem 6 states that under certain conditions, these occlusion models are flat and thus their local histograms can be analyzed according to Theorem 4. Depending on the nature of Φ and Ψ, the resulting occlusion model from their expansion may correspond at some level to textures of interest. For example, the texture of Figure 4.5(c) is somewhat akin to that of cartilage shown in Figure 4.1(a). Stronger similarities require that the real texture's shapes and colors be reasonably approximated by a probability distribution which is not always a valid assumption in all applications. However, our focus is on histopathology

applications where such assumptions are valid for many textures due to their biological origins, and thus, provide some theoretical justification for their analysis by local histograms.

A critical shortcoming of the expansion operator is that it can only generate binary occlusion models, those models formed using only two occlusion models. Such models can only describe textures that are basically formed using two colors, i.e., foreground and background. To model even more complex textures, we propose our next occlusion model generating operator that *lays* one binary occlusion model on top of another in order to generate multi-valued occlusion models. Specifically, for any $\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi})$, $\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi})$ and $\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)$, we define the *overlay* of φ over ψ with respect to σ to be $\varphi \#_\sigma \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})$,

$$(\varphi \#_\sigma \psi)(x) := \begin{cases} \varphi(x), & \sigma(x) = 0, \\ \psi(x) + N_\varphi, & \sigma(x) = 1. \end{cases} \tag{4.14}$$

Effectively, this operation is the result of punching holes in the image of φ and laying it on top of an image of ψ with the locations of the holes given by σ and the values of ψ increased by a factor of N_φ to prevent confusion with φ. Examples of this overlay operation are shown in Figure 4.5.

Consecutive overlay operations beginning with simple binary occlusion models can create very complex textures than can emulate those found in real images. The nature of the overlay operation allows us to formulate a sort of hierarchy where certain images or layers can only be laid on top of specific other layers. These hierarchies can reflect the natural spatial relationships between various components that make up a texture. For example, nuclei are generally expected only to occur within the boundaries of a cell. If we begin with two binary images where one represents the cells and the other the nuclei, a careful ordering of the overlay operation will result in a texture where the nuclei are only found within the cells. This example if very similar to that shown in Figure 4.6(c) though since cells do not contain multiple nuclei we would refine our image representing nuclei to better reflect this.

(a) σ (b) σ' (c) $0\#_\sigma\sigma'$ (d) $\sigma\#_{\sigma'}0$
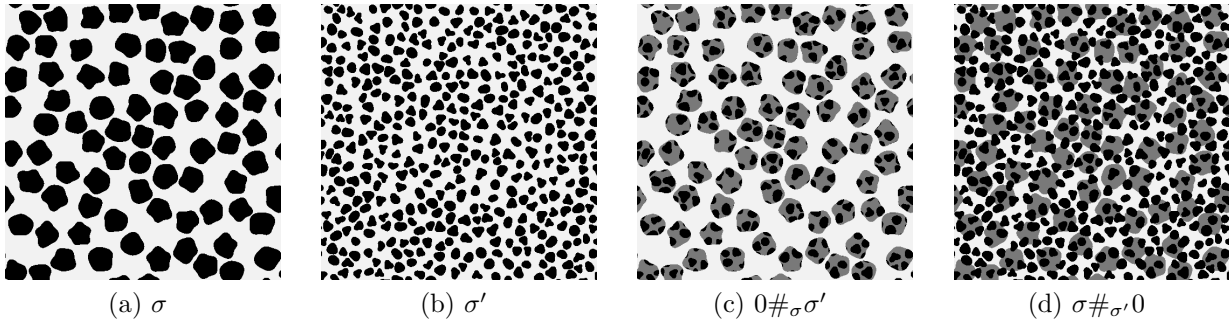
Figure 4.6: Examples of the overlay operation (4.14). For the sake of clarity, larger values are represented by darker shades. We present the two $\{0,1\}$-valued label functions $\sigma = \varphi \star \{\psi_x\}_{x\in\mathcal{X}}$ and $\sigma' = \varphi' \star \{\psi'_x\}_{x\in\mathcal{X}}$ of Figure 4.5(c) and (c) again in (a) and (b), respectively. For simplicity, consider a constant function $0 : \mathcal{X} \to \mathbb{Z}_1$ that assigns label 0 to every point in x. The result of overlaying (4.14) 0 over σ' is shown in (c) where σ-shaped holes are effectively cut out of 0 and the result is laid over σ'. A different texture can be produced by instead cutting σ'-shaped holes out from σ and laying the result over the constant function 0 as shown in (d). More applications of the overlay operation on either (c) or (d) will result in even more complex textures.

Using fixed label functions, the overlay operation produces a deterministic result but, as was with the expansion operator (4.11), using occlusion models produces a probabilistic result. Formally, given probability density functions P_Φ, P_Ψ and P_Σ on $\ell(X, \mathbb{Z}_{N_\Phi})$, $\ell(X, \mathbb{Z}_{N_\Psi})$ and $\ell(X, \mathbb{Z}_2)$, respectively, we define the overlay of the occlusion model Φ over Ψ with respect to Σ to be the new occlusion model $\Phi\#_\Sigma\Psi$ whose probability density function is $P_{\Phi\#_\Sigma\Psi} : \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \to [0, 1]$,

$$P_{\Phi\#_\Sigma\Psi}(v) := \sum_{\substack{\varphi\in\ell(\mathcal{X},\mathbb{Z}_{N_\varphi}) \\ \psi\in\ell(\mathcal{X},\mathbb{Z}_{N_\psi}) \\ \sigma\in\ell(\mathcal{X},\mathbb{Z}_2) \\ \varphi\#_\sigma\psi=v}} P_\Phi(\varphi)P_\Psi(\psi)P_\Sigma(\sigma). \tag{4.15}$$

We will shortly prove that (4.15) is a proper probability density function. Furthermore, we will also show that the corresponding occlusion model $\Phi\#_\Sigma\Psi$ is flat provided Φ, Ψ and Σ are flat. As was the case with the expansion operator, the flatness of these occlusion models means that the local histograms (4.1) of composite images (4.3) produced by such

an occlusion model will follow Theorem 4.

Theorem 7. *If Φ, Ψ and Σ are occlusion models on $\ell(X, \mathbb{Z}_{N_\Phi})$, $\ell(X, \mathbb{Z}_{N_\Psi})$ and $\ell(X, \mathbb{Z}_2)$, respectively, then (4.15) defines a probability density function on $\ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})$. Moreover, if Φ, Ψ, and Σ are flat, then $\Phi \#_\Sigma \Psi$ is flat.*

The *overlay* operation allows us to create complex occlusion models by laying simpler occlusion models on top of each other in a controlled manner. Depending on the texture of interest, particular components represented by certain occlusion models should only be present where certain other occlusion models are present. In the context of our histopathology applications, this refers to situations such as nuclei only residing within cells. Furthermore, we have shown that overlay operation applied to *flat* occlusion models will result in a more complex but still flat occlusion model. With this operation we conclude the mathematical framework's current form and now compare it to some related works.

4.4   Comparison to Related Works

Neither local histograms nor occlusion are novel concepts by themselves and have been used in many different ways for image-related applications. The novelty of our mathematical framework comes from relating the two in a non trivial and rigorous manner that supports a variety of image-classification tasks. By presenting this strict mathematical analysis, we have provided some theoretical justification for the use of local histograms for histopathology-related applications and also put forth the conditions under which they may be applied to other applications and still benefit from the theory. We now present a discussion of related works and comment on similarities and differences. As we strive to maintain a consistent terminology throughout this thesis, some terminology may differ from that found in referenced works.

4.4.1 Mathematical Morphology

Mathematical morphology refers to large body of related work that seeks to analyze shapes, most often in digital images, but often applied to other areas where the properties of geometrical structures is of importance. Pioneered by Matheron and Serra [57–59], mathematical morphology was initially formulated to aid in analysis of various materials of interest to material science. This novel practical approach leveraged knowledge from many disciplines including set theory and integral geometry, providing much of its mathematical foundation from which many advances would be made.

As work progressed, mathematical morphology was applied to binary images and resulted in the creation of morphological operations [58–60], such as erosion, dilation, opening, and closing. The formulation of these operations was done in an axiomatic fashion so as to eliminate ambiguity about the operations' intended effects. Though initially created as deterministic operations, these operations were generalized to images based on certain random models.

Additional work further extended mathematical morphology to grayscale images. This not only widened the applicability of the existing operations but also resulted in new ones such as the watershed transform [61], a commonly used segmentation algorithm. The increased applicability of mathematical morphology further increased its popularity to the point at which it is now considered a fundamental method for shape analysis in images.

One relation between mathematical morphology and our work concerning local histograms and occlusion models is that of shape analysis be it explicit or implicit. To describe shapes, mathematical morphology seeks to explicitly quantify them, while we do that implicitly to understand the identity resulting in the shape. We will not focus on comparing the vast works of mathematical morphology to our own but rather focus on two very relevant concepts that directly relate to our own. The first is a general concept called Boolean models [59, 62, 63] that have a strong resemblance to our expansion (4.11) and overlay (4.14)

operators. The second is an image formation model called the Dead Leaves Model [59,64,65] that is based on Boolean model ideas and seeks to understand various properties of natural images.

4.4.2 Boolean Models

The Boolean model is a popular model of random sets that was originally conceived by Matheron [62] for material analysis. A common situation in material analysis is that where germs are embedded in some base substance and then allowed to grow into grains. This process is often referred to as nucleation and growth in the material science community. The behavior of this process for a variety of applications is an active area of research and thus mathematically modeling it is a useful endeavor.

Matheron proposed a model where the spatial locations of the germs are modeled by a stationary Poisson process and the corresponding grains are modeled by a distribution of random sets. The random distribution of germs reflects the heterogeneity of the base substance while the random shapes of the grains reflects the behavior of the nucleation and growth process. To formally define the model, let Ω_λ be a stationary Poisson process in $\mathbb{R}^d$ with rate λ, and $\{S_{0,0}, S_{0,1}, S_{0,2}, \ldots\}$ be a series of i.i.d. random closed sets independent of Ω_λ. From these, the random closed set:

$$S = \bigcup_{x_i \in \Omega_\lambda} (S_{0,i} + x_i) \tag{4.16}$$

is said to be the Boolean model with typical grain $S_{0,0}$. Since our focus is on images, we restrict $\mathbb{R}^d$ to $\mathbb{R}^2$. The locations of the germs are governed by Ω_λ while $\{S_{0,0}, S_{0,1}, S_{0,2}, \ldots\}$ are the shapes of the grains associated with each of the germs. The typical grain $S_{0,0}$ is often the average or most representative grain shape amongst the possible variations. This definition of the Boolean model is also referred to as a single-phase Boolean model (SPBM)

since the seeds and their corresponding grains are assumed to all be of the same material or phase.

It is easy to see that the SPBM formulation is functionally identical to our expansion operation (4.11) with some minor differences. Both methods are two step processes that first randomly distribute points in a given space and then grow shapes centered on those points. In spite of their general equivalence, there are particular similarities and differences that we now discuss.

The most relevant similarity is that of the stationarity [66] of the Poisson process Ω_λ in (4.16) and the desired translation invariance of Φ as detailed in Theorem 6. Stationarity of Ω_λ means that $\Omega_\lambda(x')$ and $\Omega_\lambda(x' - x)$ are equal, i.e., that Ω_λ is translation-invariant. It is clear that while the terminology used in both formulations differs, they are equivalent. Still, as a result of the context in which the SPBM and the expansion operation were conceived, there is an important difference between the two.

This difference is the strong, if not absolute, preference of the SPBM (4.16) to use a Poisson process for the distribution of the seeds while the expansion operator (4.11) has no such preference. This choice reflects the material science focus of Matheron's intended application, the natural behavior of the material processes. Additionally, by restricting the distribution of the seeds to be a Poisson process, some properties of SPBMs may be more easily derived and proven. As a result, much of the existing and practical work related to the SPBM is highly conditional on the use of a Poisson process.

In a very similar manner to our own transition from the expansion operator (4.11) to the overlay operator (4.14), the SPBM was later extended to the multiple-phase Boolean model (MPBM) in order to model materials that consist of two or more phases. The basic formulation of the MPBM is that there are N independent phases, each modeled by a member of a set of SPBMs $\{S_i\}_{i=1}^N$, that are consecutively grown, one on top of another. Using Ω_{λ_i}, λ_i, and $\{S_{i,0}, S_{i,1}, S_{i,2}, \ldots\}$ to define S_i according to (4.16), the superposition of

indexed random closed sets defined by:

$$S\big(x \in \{S_i \setminus \bigcup_{j=1}^{i-1} S_j\}\big) = i \tag{4.17}$$

for $i = 1, \ldots, N$ is the MPBM consisting of N phases.

The phase corresponding to $i = 1$ is considered the top phase while that corresponding to $i = N$ is the bottom one which means that phase 1 is actually grown last while phase N is grown first. Thus, any point x in the overall material is considered to belong to phase i if that phase is visible at that location.

As was with the SPBM and our expansion operator, a strong similarity exists between the MPBM and the overlay operator. Both methods are ordered superpositions of random sets. In the case of MPBMs and material analysis, the superposition order reflects the actual growth process of the material of interest. For our histopathology application, the order of consecutive overlay operations models the spatial relationships between the different biological components. Both are functionally very similar but the intuition behind the order of superposition differs.

In spite of some functional similarity, a major difference between these two formulations exists. The MPBM (4.17) is merely the ordered superposition of random sets which means that phases grown later will cover any phases grown earlier. As a result, each phase is independent of other phases with no explicit interaction between them, i.e., the configuration of any given phase does not influence any other phase. In contrast, the overlay operation (4.14) is the ordered superposition of occlusion models where those laid down later do not necessarily occlude those laid down earlier. The overlay operation allows us to choose which of the occlusion models laid down prior to a given occlusion model will be occluded. This control is critical to correctly modeling images in certain applications such as our histopathology ones.

It would be naive to claim that the MPBM could not be modified to reflect similar rela-

tionships between phases as exists between occlusion models in the overlay operation. However, much existing work derived from MPBMs utilizes the straightforward superposition of phases to simplify and strengthen many conclusions made. Furthermore, our formulation of the overlay operator demonstrates our focus in creating a mathematical framework that reflects the spatial relationships between different components represented by occlusion models.

Though a variety of work related to analysis of Boolean models exists, we will not compare our work to all of them. Rather, we emphasize that our mathematical framework focuses on local histograms and how they relate to our occlusion models. Our framework was conceived and developed with the purpose of aiding classification of certain families of textures. The work derived from mathematical morphology is much more general and thus not necessarily as powerful in many applications.

As both methodologies are strongly grounded in rigorous mathematical analysis, it is mostly a difference in language and intended practical application that separates them. Major differences in basic assumptions such as the strong prefer

4.4.3   Dead Leaves Model

Matheron and Serra's pioneering work on mathematical morphology allowed for many useful theories, algorithms, and methods of analysis to be developed for a variety of areas. One such area is the modeling of natural images, those images seen by the naked eye. As was discussed in Section 2.2, automation of everyday visual tasks, many of which utilize natural imagery, is a highly active area of research and development.

The modeling of natural images is a relatively rich area of research though there are many distinct goals and approaches. In general, such models want to describe natural images in a way that makes extraction of useful statistics from them easier. These statistics can then be used for a variety of applications, two of the most popular being noise removal and

understanding the effect of scale. We will not discuss these topics in detail but instead focus on one particular model of natural images that is related to our own mathematical framework and refer readers to [67] for a review of natural image modeling.

This model is the Dead Leaves Model (DLM), a specific version of the MPBM discussed earlier. Attributed to Matheron [64] and Serra [59], the DLM distinguishes itself from the basic MPBM by making certain assumptions about grain shapes and their associated distributions. The DLM assumes that all grains are the same basic shape but their location, scale, rotation, and appearance are governed by some random distribution. For example, DLMs commonly use single color circles where the scale (radii) and color of any given circle is determined by a random distribution. Other forms include ellipses and rectangles with their shape properties being randomly distributed as well.

Though the DLM has been known for some time in a variety of forms, only relatively recently have researchers presented a thorough study of statistics of natural images modeled by a DLM. In particular, Lee et. al [65] have performed a thorough statistical study of natural images utilizing the DLM. The primary goal of this work was to determine if a simple version of the DLM could realistically model various useful statistics of natural images. Rather than attempt to detail all the various experiments and conclusions that were drawn by the authors, we present only the main conclusions and refer readers to the actual work for more information.

The primary conclusions reached in this work was that simple DLMs are capable of modeling a variety of statistics for natural images. These statistics include pixel intensity difference distributions, scale specific pixel intensity difference covariances, and Haar wavelet coefficient distributions. Furthermore, the accuracy of these models was far from trivial and strongly support the observation that natural image statistics are highly non-Gaussian. From these observations, the authors showed that DLMs have strong practical applicability in natural image problems. Naturally, our interest in the DLM is how closely it relates

to our own mathematical framework. In addition to the similarities and differences of our own framework to the MPBM, an additional and significant difference exists between our framework and the DLM.

This difference concerns the method of description of the images used in each approach. In the DLM, images are described through various statistics while in our framework they are described by local histograms. Though local histograms are still a statistic in the truest sense, our use of occlusion models to relate those of simpler images to those of more complex images (Theorems 3 and 4) does not have a parallel in the DLM analysis. Rather, in the DLM analysis, these statistics are taken to represent the image alone. Though the construction of images by the DLM is done by the superposition of various independent grains, only the cumulative nature of the grains is considered. The extension of the statistics used in the DLM analysis to mimic our own local histogram approach can be done but we have not encountered such work in the literature.

Additionally, the authors acknowledge areas of improvement for the DLM [65]: "We have furthermore found evidence that the current version of the dead leaves model can be further improved by (1) using suitable primitives, (2) adding textures to the primitives, and (3) taking the hierarchical structure of objects into account." Our interpretation of these statements is that our mathematical framework actually accomplishes all of these extensions. The expansion operator (4.11) addresses the first improvement by placing no explicit restriction on the primitive or blob used to construct the overall occlusion model. The second improvement is accomplished through our formulation of occlusion (4.3) since we occlude images as opposed to random sets. Each image can have distinct and complex textures though this may weaken the overall result of Theorem 4 in comparison to those images composed of simple textures. The final improvement is achieved through our overlay operation (4.14) as discussed previously. In spite of the fact that no related work we have encountered addresses all these additional improvements, we acknowledge that extensions of

the DLM to incorporate these items is feasible.

Chapter 5

Classification System

We now present the second main contribution of this thesis: a classification system for the identification and delineation of tissues and pathologies in H&E-stained images based on expert domain knowledge. We begin by detailing our methodology for expert domain knowledge based feature design called the HV, and follow with a discussion of the purpose and implementation details of the derived features. We then present our approach to achieving both identification and delineation via *pixel-level classification*. We conclude by discussing the challenging task of rejection that is needed in certain applications. While Chapter 4 presented the primary theoretical contributions of this thesis, this chapter aims to describe the algorithmic and implementation related aspects of the overall work.

5.1   Histopathology Vocabulary

Histopathology classification problems are expert classification problems (see Section 2.3); the pathologist's extensive training and experience that allow him to perform very complex histopathology tasks is clear evidence of this. Almost every effort in designing machine learning tools (including classification systems) for histopathology is based on some task that a pathologist is already capable of. Many of these seek to automate these tasks for the

sake of improved efficacy and efficiency.

The pathologist's participation is critical to the design of these systems by allowing the engineer to leverage the his knowledge and experience. For classification systems, feature design is likely to be initially based on the pathologist's input such as application appropriate feature types. In interpreting this input, the engineer must address a critical challenge: his lack of a detailed pathology knowledge.

The pathologist and the engineer each use specialized languages. Specifically, the pathologist's medical language differs greatly from the engineer's computational one. There must be enough communication between them to ensure that the engineer's interpretation of the pathologist's input is faithful to the pathologist's knowledge. However, for many works, not only is the engineer's interpretation not necessarily verified, but it may also be too specialized to the task at hand.

As stated in Section 1.4, a single pathologist is capable of almost all the tasks automation efforts address without any explicit retraining. This implies that there is some fundamental collection of knowledge that is used across applications. While some applications may require extremely specific knowledge, the majority of them are well within the grasp of a reasonably trained and experienced pathologist. Therefore, in our efforts to automate various histopathology tasks, we must ask if it is better to address each task individually or to address many tasks simultaneously as the pathologist does?

One could argue that by focusing on an individual application, automation of the task becomes more feasible. However, the algorithm for a specific application is most likely ill-suited for an unrelated application. Furthermore, very few algorithms achieve levels of performance that match the pathologist in spite of their single application focus. For these reasons, we believe that a larger impact can be made by learning from the pathologist's knowledge in a verifiable manner and by addressing many applications with the same framework.

Thus, we propose the HV as an expert-guided feature set created through the collabora-

tion of pathologists and engineers, using strong feedback mechanisms to ensure effectiveness. The HV will be an intuitive feature set that reflects the pathologist's knowledge and specifically targets those discriminative qualities used by him. By basing its derivation on the extremely general histopathology application of teratomas discussed in Section 2.1.1, we will create a feature set that can be used for many applications without significant modification. Naturally, the idea of leveraging expert knowledge is not novel by itself but by proposing a formal methodology, we will avoid the pitfall of not verifying the engineer's interpretation of the pathologist's input. We will first discuss and illustrate the basic methodology and then explain the feature set derived by applying it to histopathology.

5.1.1 Methodology

We start by describing the basic methodology used to formulate the HV, shown pictorially in Figure 5.1. The primary goal of this methodology is to create a set of computational descriptors/features that are derived from the specialized knowledge of the expert. This approach uses iterative communication between expert and engineer to ensure that both parties understand each other as much as possible. We now describe each of the 5 main steps of the methodology in the context of histopathology.

Formulation of Initial Set of Descriptions

For almost every histopathology classification system, the pathologist serves as the ideal existing system whose method and performance sets the standard for the engineer. When addressing tasks such as identifying tissues and pathologies or grading the progression of a disease, a pathologist can provide a wealth of information. Thus, the first step of the methodology is for the pathologist to describe how he identifies/discriminates each tissue, pathology, or other label of interest.

As discussed earlier, without any modifications, the pathologist's descriptions are likely to
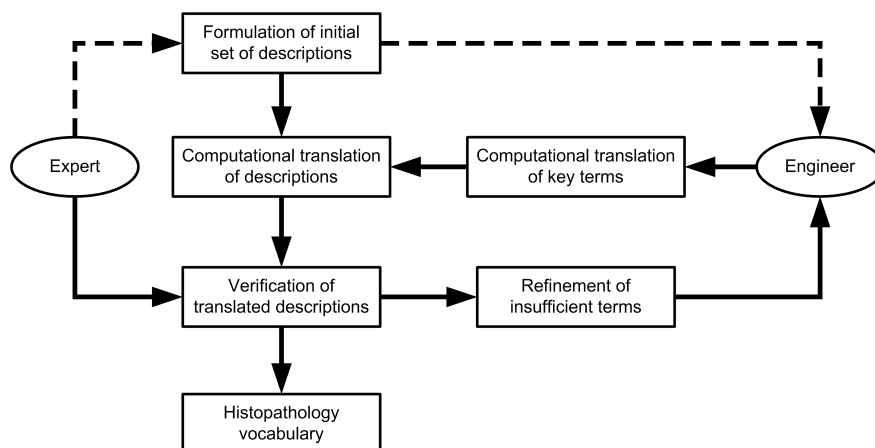
Figure 5.1: Flowchart demonstrating basic HV formulation methodology. Dashed lines indicate one-time steps while solid lines are part of the iterative/feedback portion of the process.

use medical terminology that the engineer does not understand. To avoid this, the pathologist is asked to provide these initial descriptions using as simple and accessible language as possible. This often corresponds with avoiding the use of medical language and dividing complex concepts into collections of simpler ones.

These descriptions should also reflect the importance of certain descriptive characteristics for each label of interest. For example, when identifying bone, the background material color is the most important characteristic while for fat it is the presence of lumen. The ranking of these characteristics allows the engineer to better understand the pathologist's method for identifying each label. Later we will show how these ranked lists describing each individual label are combined to provide a global list describing all the labels together.

This step in some form or another is performed during the majority of collaborations between pathologists and engineers. In such applications, an engineer should not design a classification system based purely on a problem definition and his own knowledge. However, in most works, this is the limit of communication between pathologist and engineer before any actual algorithm design takes place. Later steps of the methodology will address this shortcoming but we now detail the step of converting the pathologist's descriptions into

computational terms.

Computational Translation of Key Characteristics

The pathologist's lists of ranked descriptive characteristics for each label are used to form a set of computational counterparts or synonyms from which features can be implemented. If the engineer is to use the pathologist's input in designing features, he must determine how to computationally model them. For many works, this computational implementation happens immediately without any sort of validation of how correct the engineer's interpretation is, a potential mistake that we wish to avoid. To perform this verification, the engineer must give some indication as to how he might implement the features based on the pathologist's descriptions, i.e., a computational translation.

This computational translation is simply the determination of computational synonyms for the various descriptive characteristics put forth by the pathologist. Computational synonyms are terms that are understandable to the non-engineer while still having a clear computational meaning or implementation. For example, a description of color can be implemented using any one of a number of colorspaces. In contrast, a description of texture can have many interpretations that may not correspond to what the pathologist is actually describing. Furthermore, the computational translations must still be reasonably well understood by the pathologist so that he may verify them. For example, the pathologist's description of "box-shaped tissue" by the engineer into "shape whose chain code has 4 distinct runs and the difference between changes is 2" is more than likely not well understood by the pathologist.

After the computational translation, the engineer distills the most globally important characteristics based on their relative importance to each individual label. By doing so, the engineer begins to form the feature set best suited to describing all the labels collectively. It is from this process that the methodology's title of "vocabulary" is motivated. The

pathologist's descriptions represent a collection of category-value pairs where a category is a particular descriptive characteristic type and a value is a specific member of this category. For example, color and shape are categories while red/blue/white and circle/square/triangle are example values, respectively.

The first iteration of this step results in a first draft of the computational vocabulary that will be used as the blueprint for the implemented features. The engineer has distilled from the pathologist's collection of category-value pairs, the categories that are the most discriminative and can be computationally modeled. The next step enters an iterative feedback loop that will refine this vocabulary to maximize practical effectiveness and understanding by both pathologist and engineer.

Computational Translation of Descriptions

The engineer now presents to the pathologist the computational translations of his original descriptions. For example, the pathologist's description of striated muscle may be: "Muscle sheets are striated and parallel to each other with nuclei elongated and parallel to each other." This can be computationally translated to: "Muscle sheets are linear in shape with common orientation and nuclei are strongly elliptical in shape with common orientation." In this example, the main categories are the shape and relative orientation of muscle sheets and nuclei. Using these computational translations, the pathologist will now either validate the engineer's interpretation or identify areas of improvement.

Verification of Translated Descriptions

In this step, the pathologist checks the computational translations of his original descriptions for correctness. As was stated earlier, the engineer's lack of a formal pathology knowledge may lead to misinterpretations of the pathologist's descriptions. By inspecting the computational translations of his descriptions, the pathologist will determine whether or not his

knowledge has been faithfully modeled.

The pathologist performs this verification by seeing if he can identify the label being described based on the computational translation alone. In a way, this verification emulates the classification process where the translated descriptions are the features and the pathologist is the classifier. If the pathologist is able to correctly identify the label and all aspects of the translation are reasonably well understood, then we have some validation that the translation is sufficient to describe the label. However, should the translation not insufficient in some way (either leading to an incorrect label or having unclear aspects) then it must be refined.

For example, the translation for cartilage may be: "Of the tissue's average RGB value, the B value is significantly greater than that of R. The tissue contains, on average, circularly shaped regions of lumen whose spatial density ranges from dense to sparse. Within these lumen regions, nuclei can be found whose average B value is overwhelmingly greater than the R value." The pathologist will identify this as describing cartilage though he may be uncertain about certain aspects of the description such as the how dense and sparse compare to each other. To address these insufficiencies, the overall translations are refined by refining the computational vocabulary term.

Refinement of Insufficient Terms

We refine the computational vocabulary so that it is more effective in describing labels, more faithful to the pathologist's knowledge, and more understandable by both pathologist and engineer. In the previous step, the pathologist identified aspects of the computational vocabulary that were either insufficient in either discriminative power or clarity and must now be refined by the engineer. For example, assume that the current computational vocabulary contains a term for the collective nuclei orientation in a local area. Specifically, the entropy of the discrete histogram of the individual nuclei orientations in a local area. For tissues

whose nuclei only have a few orientations, the entropy will be small, while for tissues whose nuclei are randomly oriented, the entropy will be large.

Therefore, for cartilage, whose distribution of nuclei orientations is close to random, this term would be: "The local nuclei orientation histogram has high entropy." However, this description is probably not well understood, if at all, by the pathologist. Therefore, the engineer must refine this term so it is better understood by the pathologist while still having a clear computational meaning. For example, a possible refinement is: "The average difference in orientation between all pairs of nuclei in a local area." Based on this refinement, the translation is updated to: "The average difference in orientation of nuclei is greater than 45°." This is most likely to be better understood by the pathologist and when combined with other terms, allow the pathologist to identify cartilage.

After the current round of refinement, the translations are updated and then presented for verification. This iterative feedback loop persists until no further refinement is required. It may also be that there are terms that are too abstract or specialized to have computational translation that is well understood by both pathologist and engineer. Those terms that require no refinement have demonstrated sufficient discriminative power, faithfulness to the pathologist's knowledge, and clarity for both parties to be added to the HV feature set.

Histopathology Vocabulary

The iterative feedback loop formed by the previous steps has refined the computational vocabulary terms to a level where we can confidently create features based on them. By maximizing the coherence between the pathologist's domain knowledge and the engineer's computational interpretation, we increase our understanding of the algorithm's behavior. Therefore, we have a justified belief that the implemented features are practically effective and faithful to the pathologist's knowledge. In this work, we apply this methodology to our teratoma application (see Section 2.1.1) whose resulting feature set we discuss next.

5.1.2 Vocabulary

Using the proposed methodology, we have created an initial HV consisting of the 8 features summarized in Table 5.1. To utilize our mathematical framework and support our delineation goal, these features are computed in a local neighborhood around a given pixel (see Figure 5.2 for some examples). By computing local features, each pixel provides a local description of its tissue. In this section we describe the HV features and their implementations. We begin with our algorithm for the segmentation of important tissues components: cytoplasm, lumen, nuclei, and background material.

1. Background/fiber color	5. Nuclei density
2. Cytoplasm color	6. Nuclei shape
3. Clear areas (lumen)	7. Nuclei orientation
4. Nuclei color	8. Nuclei organization

Table 5.1: HV features ranked according to their discriminative power with respect to teratoma tissue identification.

Component Segmentation Algorithm

An important part of the extraction of many of the HV features is the segmentation of components that are common to the majority of tissues and pathologies. Cytoplasm, lumen, nuclei, and background material are amongst the most important of these and in this section we describe an algorithm for their segmentation. Segmentation is a well studied problem and we do not focus of the development of a novel approach, but rather a simple one that works reasonably well under a variety of conditions.

The nature of H&E staining, as discussed in Section 1.1, supports a segmentation approach based on analyzing the differing contributions of H&E stains to each component. The general appearance of a component is the direct result of absorbing a certain amount and ratio of H&E resulting in some combination of blue-purple and pink-red coloration. For

our work, cytoplasm is primarily pink, lumen is white, nuclei is dark-blue, and background material is a mixture of colors.

The difficulty in this segmentation arises from the variations that contribute to a component's exact appearance (see Chapter 1). Specifically, the exact shades of the components colors described above can vary in a non-negligible way, something the segmentation algorithm must account for in order to be effective and widely applicable. We propose to normalize our data by using a demixing/stain separation algorithm for H&E-stained images proposed by Macenko et. al [25].

Demixing algorithms in image analysis try to decompose complex images into combinations of simpler parts. For example, a standard RGB image can be demixed into linear combinations of pure red, pure green, and pure blue vectors, i.e., the individual channels of the image. The intuition for using demixing for H&E-stained image normalization is that the image's appearance is the result of absorbing H&E stains in specific ratios. The variations that occur are the result of these ratios being altered. If we make some assumptions about how these ratios are altered, we can normalize them across images and thus account for the corresponding variations.

The method we use [25] utilizes a specific color representation and the SVD to demix H&E-stained images into images representing each stain. Figure 5.3 presents some examples of this demixing or stain separation. This algorithm not only determines the relative amounts of the H&E stains that is absorbed at each pixel, but also achieves the desired lumen segmentation. For details, we refer readers to the work itself.

The segmentation of the remaining components is based on the assumption that they are present in any given image and that regardless of the exact nature of the variations in an image, the variations affect the entire sample consistently, i.e., affect each component equally. While only an assumption, our experimentation and knowledge of the H&E staining process supports this at a reasonable level. Based on this assumption and our lumen segmentation,

we segment cytoplasm and nuclei with the remaining portion of the image being labeled as background material.

The first step is to transform the standard RGB image to the L*a*b* colorspace. In this space, the luminance of the image is captured by the L* channel and thus can be ignored to help mitigate illumination variation. Using the remaining channels of a* and b*, we apply a simple K-means clustering [44] to compute clusters representing different shades and combinations of the blue-purple and red-pink coloration of H&E, respectively. Based on experimentation, we determined that $K = 5$ clusters was fairly robust in identifying different color and shades. Our intuition is that two clusters will represent shades of hematoxylin's blue-purple with another two representing shades of eosin's red-pink. The remaining cluster will collect unclear or ill-defined combinations of these stains.

Since K-means incorporates some randomness in it's overall algorithm, we must determine the correspondence between clusters and colors/shades. We apply heuristics using the b* component and the desired constancy in the appearance of the components. The b* component represents a color's position between yellow and blue. Therefore, a reasonably robust way to order the clusters in terms of their color is according to the value of the cluster center's b* component. We now have two pairs of clusters, one that is predominantly blue-purple while the other is red-pink, leaving a final cluster of indeterminate color.

However, the average color of a cluster does not reflect its constancy within that cluster. Thus, we apply a second ordering based on the average distance of a cluster's members to its center as an estimate of the cluster's compactness. Each of the two pairs of clusters determined previously is further ordered based on this compactness measure. The cluster that is more compact is labeled as representing one of the desired components, nuclei for the blue-purple clusters and cytoplasm for the red-pink clusters. The pixels corresponding to these clusters are now labeled as one of these components though some additional refinement is performed to reduce errors and noise.

A key consideration is that these components are spatially connected objects and not just random collections of pixels. Labeling completely isolated pixels or small and irregularly shaped collections of pixels with the same label are most likely errors. The segmentation has so far ignored any spatial information and is likely to make such errors. To address this, we apply a series of morphological operations to each of the labeled components to first remove connected groups of pixels smaller than a specified size appropriate for the current magnification. Furthermore, for nuclei which are generally elliptical, an opening operation with an appropriately sized circle removes spurs and connections between adjacent nuclei. After this refinement, the labels for cytoplasm, nuclei, and lumen have been finalized with the remaining portion of the image being labeled as background material.

HV Features

In this section we describe the implementation of each of the 8 HV features detailed in Table 5.1. As many of these features have parallels in other works, our implementation reflects both our mathematical framework and the local nature of the features. Naturally, we expect future work to not only refine the existing features but also develop new ones. We begin this section with a discussion of some aspects that are common to many of the features.

Implementation via local histograms is our preferred approach but due to the computational constraints, it is not always done. For example, statistics such as means and moments can be extracted from the appropriate local histograms, but when attempting to compute them for large images, the memory requirements become unrealistic and we must avoid explicit computation. For example, the local average color of an image corresponds to the sum of the bin positions, each weighted by their value, of the color local histogram. Therefore, we present our current implementations where we often forgo explicit local histogram computation in order to achieve reasonable computational efficiency. As hardware

and algorithms improve, we hope to implement as many features as possible using local histograms to strengthen the connection between algorithm and theory.

Locality or scale is a notion that we discussed in Chapter 4. For our identification and delineation goal, we classify each pixel individually. However, using only that pixel's information will not yield desired or robust performance. Thus, we classify using both the pixel's and its neighbors' information (see Chapter 1). Naturally, the choice of these neighbors has a major impact on the overall performance of the algorithm. We use a 2-D filter whose values specify the neighborhood's spatial span and the relative importance of any given neighbor. This is effectively the same approach taken in Chapter 4 to define the locality of local histograms. We design and experiment with many types of filters based experience and intuition. In Chapter 6 we discuss these experiments and present our conclusions concerning locality and its effect on overall performance.

Color representation is critical since color is the primary descriptor of so many aspects of a tissue or pathology. A very active area of research for a variety of applications, the needs of the application determine which representations are best. In this work, we seek a representation that is best suited to our classification task. Our investigation led us to explore standard colorspaces such as RGB, L*a*b*, HSV, and YCbCr.

Our experimentation sought the colorspace that provides the most consistent presentation of tissues and pathologies. In other words, given a set of examples of a given tissue or pathology, which colorspace would most tightly cluster those examples and similarly for other tissues? These experiments showed that the RGB and L*a*b* colorspaces outperformed the others in this regard without a clear winner between them. Thus, in the interest of simplicity, we have chosen to use the standard RGB colorspace throughout our algorithm except for when explicitly stated otherwise.

Background/fiber color characterizes the collective color of the tissue aside from nuclei, lumen, and cytoplasm. As this component often represents a large portion of any given

tissue, its appearance is highly indicative of identity. General descriptors such as color have been observed to be reasonably effective and robust in describing this component. Given a segmentation of the background material and our local feature approach, the primary design choice is the specification of locality via a filter.

For this feature, the main constraints for filter design are that it be rotationally symmetric and be a proper averaging filter (only positive values and sums to 1). By being rotationally symmetric we are account for the fact we do not know the exact orientation of the sample. As a proper averaging filter, the feature will extract actual colors and thus allow us to visualize them. For each pixel labeled as background material, we compute the average local color by applying the specified filter to each channel of the RGB image. For illustration we present the mathematical details of computing this feature in using local histograms as opposed to the more computationally direct method above.

The computed background/fiber boolean mask $m(x)$ specifies the positions x which will be ignored in computing the local background/fiber color. This mask along with a specified weighting function/filter w (following our design constraints stated earlier) allows us to compute the local histogram of the RGB image $f(x)$ for all possible 8-bit RGB values $y \in \mathcal{Y} = \mathbb{Z}_{256}^3$ according to:

$$(\mathrm{LH}_w f)(x, y) := \sum_{x' \in \mathcal{X}} w(x') \delta_y(f(x + x')) m(x + x'), \tag{5.1}$$

where

$$m(x) = \begin{cases} 1, & x \in background, \\ 0, & x \notin background, \end{cases} \tag{5.2}$$

The average background/fiber color $(\overline{C}_w f)(x)$ of $f(x)$ is approximately equal to the sum of the image's local histogram bin positions (RGB values) y weighted by their respective bin values values $(\mathrm{LH}_w f)(x, y)$:

$$(\overline{C}_w f)(x) = \sum_{y \in \mathcal{Y}} (\text{LH}_w f)(x, y)y \tag{5.3}$$

As stated in Chapter 4, a shortcoming of local histograms is the memory needed for storage. As a result, direct computation of (5.3) can be prohibitive due to the memory requirements of $(\text{LH}_w f)(x, y)$. To mitigate this issue, we compute $(\text{LH}_w f)(x, y)$ sequentially with respect to y and update $(\overline{C}_w f)(x)$ according to the following algorithm:

Algorithm 1 Computation of background/fiber color via local histograms
---
$(\overline{C}_w f)(x) = 0$ for all $x \in \mathcal{X}$
for $y' \in \mathcal{Y} = \mathbb{Z}_{256}^3$ **do**
  Compute: $(\text{LH}_w f)(x) = (\text{LH}_w f)(x, y')$ according to (5.1) for all $x \in \mathcal{X}$
  Update: $(\overline{C}_w f)(x) = (\overline{C}_w f)(x) + ((\text{LH}_w f)(x)y')$ for all $x \in \mathcal{X}$
end for

The final result of this algorithm is the desired weighted sum of the local histogram bin positions by their values as specified in (5.3) which represents the background/fiber color in the local neighborhood specified by w.

Cytoplasm color is effectively the same feature as the previous one though applied to a different component. Cytoplasm is a major component of many tissues, providing much of structural matrix that supports most tissues. It's almost constant presence means that differences in its appearance are good indicators of a tissue's identity. For many related tissues, such as types of muscle, the color of cytoplasm can often distinguish between them in spite of many other similarities. In designing this color feature, the same considerations of filter design and color representation are also addressed and will be discussed in the next chapter. The binary labeling from the cytoplasm segmentation is filtered with the specified filter to compute this local feature.

Clear areas (lumen) are those regions where the H&E stains are not absorbed either due to no nucleic acid or proteins being present, or a complete lack of tissue. For tissues such as fat, cartilage, and skin, the local density of lumen is a key descriptive characteristic. For

example, fat has a very high local lumen density while it is less dense for cartilage and skin. Similarly, certain tissues sometimes form around lumen during the process of the creating organs; for example, GI tissue forming into a colon.

We apply a standard filtering approach to the binary labeling of the lumen segmentation we previously computed where values of 1 represents lumen and 0 non-lumen. The result of filtering is that each pixel provides an estimate of the amount of lumen in its neighborhood. The filter is constrained to sum to 1 so that the estimate is a weighted percentage of the lumen pixels in the neighborhood. It is also constrained to be rotationally symmetric so that orientation of the sample is not a factor.

Nuclei color, while primarily blue-purple, does vary according to tissue in terms of shade and saturation, a direct consequence of different levels of nucleic acid being present. Tissue identity, tissue maturity, and whether or not cells are growing or dying will dictate how much nucleic acid is present. After specifying a rotationally symmetric filter that sums to 1, we filter the RGB image masked by the nuclei segmentation in the same fashion as we did when computing background/fiber color.

Nuclei density or cellularity is highly indicative of tissue and pathology identity. Nuclei density is highly variable between tissues making it a good indicator of them. Furthermore, given a set of related pathologies such as those concerning a particular tissue, changes in cellularity often correlates very strongly with specific pathologies. For example, as a reaction to an infection, cellularity will increase reflecting the body's immune response. This feature is computed in the same fashion as our estimation of clear area (lumen) density using the nuclei segmentation result and a specified filter.

Nuclei shape often correlates strongly with nuclei orientation and organization that collectively correlate with tissue identity. These three features partially reflect tissue and pathology structure. For example, many infections and diseases directly affect the nuclei of tissues with extreme cases completely distorting their shape. The general shape of nuclei

and regularity of that shape can very indicative of tissue and pathology.

We compute the average local nuclei shape using basic blob detection/scale-space filtering concepts [68]. Using our nuclei segmentation, we apply a series of Gaussian filters that vary in orientation and scale. This series of filters estimate the apparent size and shape of the nuclei. The collective results of the filters are refined by using non-maxima suppression to determine at each pixel which filter yielded the best match. Those determined nuclei shapes and sizes are assigned to the positions of the apparent nuclei centers. Using the specified filter, this sparse image is filtered to yield an estimate of the average shape and size of the nuclei in the neighborhood of any given pixel.

Nuclei orientation is typically a strong indicator of structure and thus identity. In particular, the orientation of individual nuclei with respect to their neighbors can tell a great deal about overall organization. It is the uniformity or randomness in local nuclei orientation that can serve as good feature. Using our nuclei segmentation, we apply a simple orientation filtering to the binary nuclei labeling. Orientation filtering most commonly refers to algorithms which use various derivatives of Gaussian and other similar filters to determine the orientation of edges in an image. These filters vary in orientation and scale in order to detect objects of varying orientations and scales.

Using our previously computed nuclei shape, we have an estimate of each nuclei's center. Thus, the orientation of the filter which yields the largest response for any given nuclei center is chosen as orientation of that nuclei. We now have a sparse image whose non-zero values indicate the orientation of nuclei. A specified filter is applied to the sparse image to compute an estimate of the average nuclei orientation in a neighborhood. The computed average is then subtracted from the sparse image, the result squared, and then filtered with the same filter. This effectively computes the variance of nuclei orientation in the specified neighborhood. The intuition is that when large numbers of nuclei have the same orientation, the variance will be small and when they have many different orientations, the variance will

be large.

Nuclei organization is a fairly ambiguous term by itself whose uncertainty is partially expressed by its lower ranking amongst the HV features. In spite of this, we want to quantify it by determining the degree of spatial isotropy that the nuclei have. For example, in skin tissue, nuclei are arranged in smoothly changing collinear layers of common orientation. Similarly, some structures specific to various tissues are formed by collections of nuclei in well defined organizations. To capture this behavior, we use our nuclei segmentation to first compute local centers of mass. These centers of mass are merely the weighted average position of the nuclei in a neighborhood around a given pixel. For each pixel, the average and variance of the distance and angle of the local nuclei to the local centers of mass is computed via filtering.

Given well structured objects, the variances will be relatively small as opposed to the large variances for randomly organized nuclei. The local averages should partially quantify the scale of the organization which may help identify tissues that contain noticeably large or small organizations of nuclei. More refined approaches could be taken to quantify this organization, but the simplicity of this approach offers some robustness to the errors that may arise from both the original nuclei segmentation and any subsequent steps.

5.2 Pixel-Level Classification

Now that we have designed our HV features, our task is to design an appropriate classifier to label regions of H&E stained images according to our set of tissues and pathologies. For those applications where an image consists of only one tissue or pathology, the labeling of regions is equivalent to labeling the entire image. In contrast, those applications where an image consists of many tissues or pathologies, we must be able to identify and label these regions individually. One standard approach is to first segment the image and then identify each segmented region. However, for reasons discussed Chapters 1 and 2, this approach cannot

be taken in all many applications due to complexity of the segmentation task. Therefore, we propose to accomplish a simultaneous segmentation/delineation and identification of tissues and pathologies by classifying each individual pixel in an image based on its local features.

A decision about the identity of a pixel cannot be reliably made based on the pixel's presentation alone. Local information, that is, the pixel's neighbors must contribute to its labeling. This is already reflected in our formulation of the HV features; they are all scale-adaptive features that compute values based on local neighborhoods. Assuming some homogeneity in the presentation of a tissue, we expect this approach to yield relatively smooth delineations. However, junctures between tissues will lead to regions of confusion where exact delineations will be called into question. However, this same uncertainty occurs when pathologists delineate images for a variety of reasons and as such we hope to operate within a reasonable degree of uncertainty.

We propose to use one of the many well-known classifiers to assign a label to each individual pixel based on its extracted HV features. Among these classifiers are the NN [46], the SVM [45], PCA [44], LDA [44], likelihood estimators [44], ensemble methods [69, 70], classification trees [71], and many more. We have experimented extensively with these general families of classifiers to determine which are better suited to histopathology related problems than others. Intuitively, the complexity and variety of problems demands a very flexible classifier and these experiments will be detailed in the next chapter.

Since we are making a decision for each individual pixel, there is inevitably some amount of pixel-to-pixel variation that leads to discontinuities in delineations. While some of these are meaningful, others, such as when a single pixel is labeled as a different tissue than all of its neighbors who share a common label, are possibly errors. To address such errors, we apply a local refinement in the form of a weighted local voting to the initial set of labels. For many classifiers, such as a NN or a probabilistic SVM, a label $\ell(x) \in \{1, 2, \ldots, L\}$ can be accompanied by a corresponding confidence $c(x) \in (0, 1]$. Given this set of initial labels,

their confidences, and a finite weighting function $w(x)$ that sums to 1, we refine the label as follows:

$$\ell'(x) = \underset{i=1,\ldots,L}{\operatorname{argmax}} \sum_{x' \in \mathcal{X}} w(x')\ell^{(i)}(x + x')c(x + x'), \tag{5.4}$$

where

$$\ell^{(i)}(x) = \begin{cases} 1, & \ell(x) = i \\ 0, & \text{otherwise} \end{cases} \tag{5.5}$$

Confidences $c(x, y)$ as well as the weighting enforced by $w(x, y)$, reflect our intuition that more confident labels and certain neighbors are likely to be the same label as a given pixel.

5.3 Rejection

In various histopathology applications, it may be necessary to actively ignore or reject data that is unclear, of unknown identity, or is not of interest. This is particularly needed with respect to our delineation goal. If given a task where the entirety of an image does not have a clear label of interest, to accurately delineate those labels we interested in, we must reject those unwanted regions. However, rejection in any classification task is an extremely challenging task. While this is a generalization, rejection in classification tasks is the need to define the boundaries between those defined labels of interest and all other types of data, be they defined labels of non-interest or purely unknown labels. Defining a boundary of any sort is equivalent to characterizing one or both of the regions that it separates.

Regardless of the approach taken, the nature of the data to be rejected is often the most challenging aspect of the task. Given a set of defined labels, the task of classifying some subset of them and rejecting the complementary subset is simply another classification task. In such a task, we merely have to classify all of the defined labels and subsequently reject those designated for rejection. However, if we are given such a set of defined labels in addition to undefined set of data, the task becomes more difficult.

The challenge in such a situation is learning both the structured nature of the defined labels to be rejected and the unstructured nature of the undefined data. In our histopathology tasks, structured components come from well-defined tissues and pathologies that are of non-interest, while the unstructured components come from tissues that are not yet well-defined and completely undefined tissues. This is of particular relevance to our teratoma application where the maturity of tissues and pathologies vary along with tissue regions that have no identity, i.e., background/noise.

To address this task, we investigated two primary approaches: rejection based on perceived confidence and rejection via classification. The first approach merely rejects labels provided by the algorithm if the associated confidence does not meet some criteria, e.g., a threshold. The confidence associated with each label is a function of the classifier we use; some classifiers provide pure probabilistic measures while others must be further manipulated to resemble a confidence. The second approach uses a wide variety of classifiers that directly address the rejection task as part of their design. We will discuss the particular methods we investigated for both of these approaches, along with our results and conclusions on rejection, in Chapter 6.
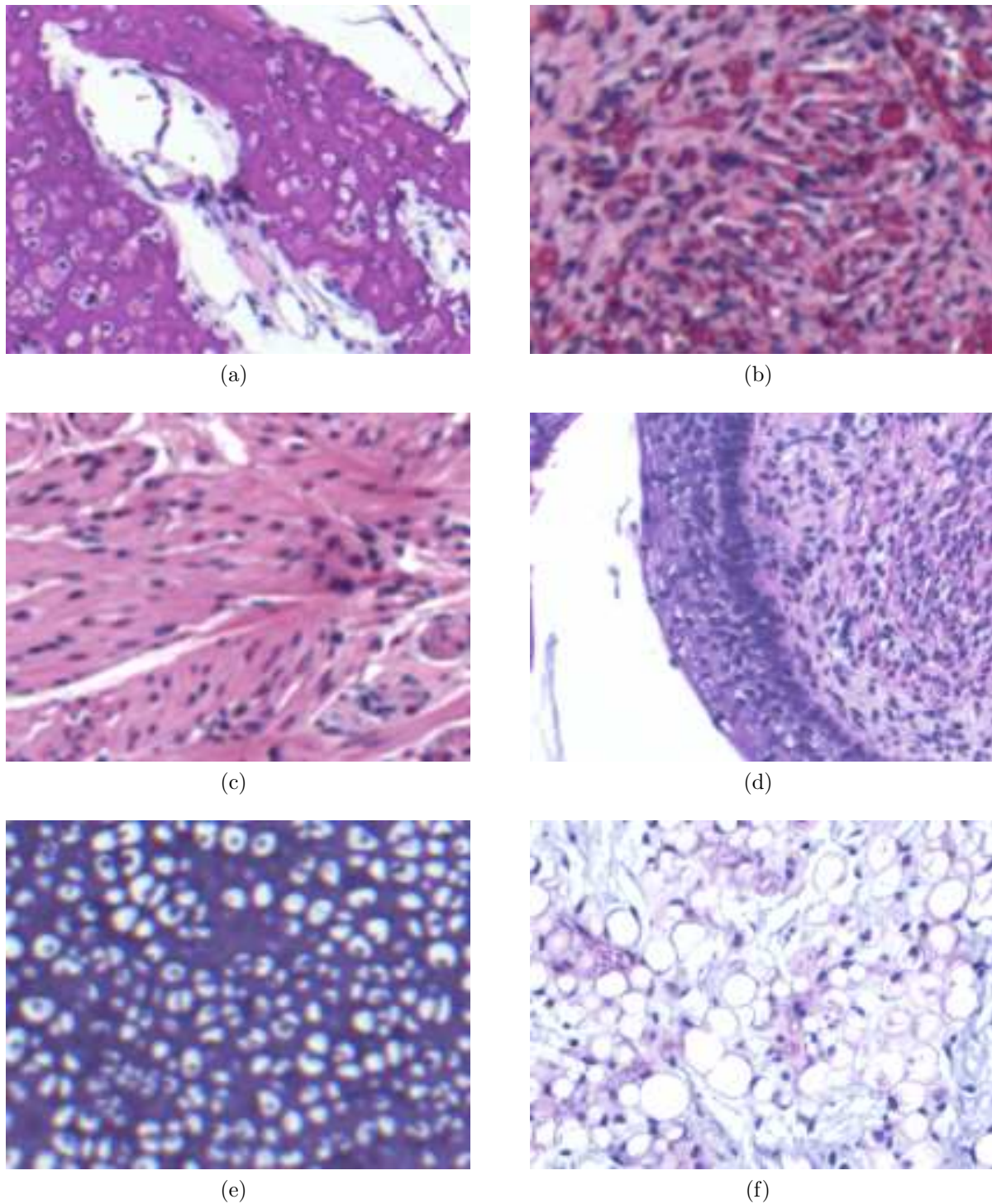
Figure 5.2: The appearance of some HV terms in different tissues. Background/fiber color for (a) bone is bright pink, while it is (b) dark pink for pseudovascular tissue. Nuclei orientation and organization for (c) smooth muscle is unstructured, while it is (d) structured for skin tissue. Clear areas (lumen) for (e) cartilage are less dense than for (f) fat.
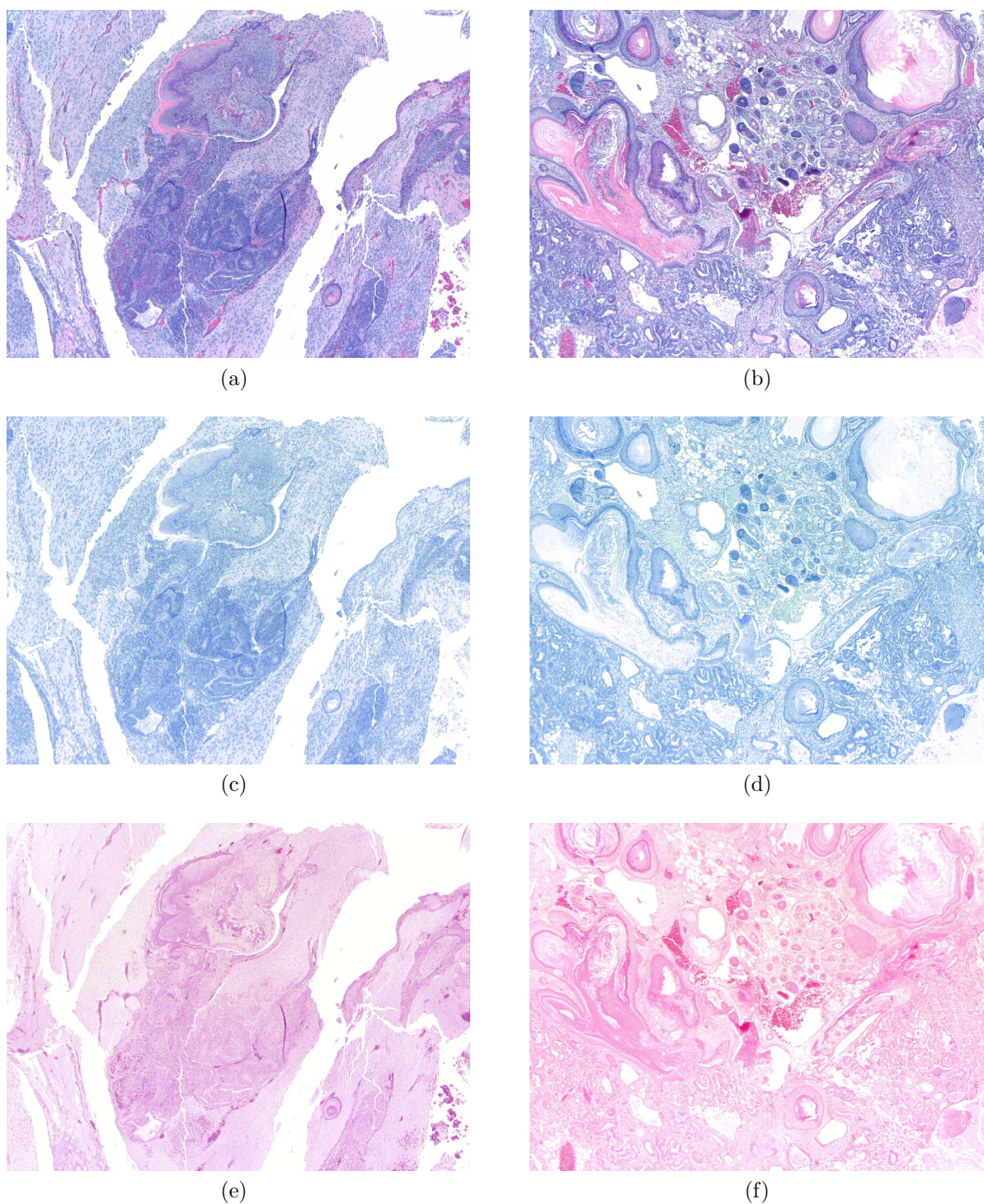
Figure 5.3:   Examples of H&E stain separation. The first row presents the original H&E stained image with the second and third rows showing the computed hematoxylin-only and eosin-only separated images. Take note of how structures primarily saturated with one stain in comparison to another appear more dominantly in the corresponding stain separated image. Continued in subsequent figure.
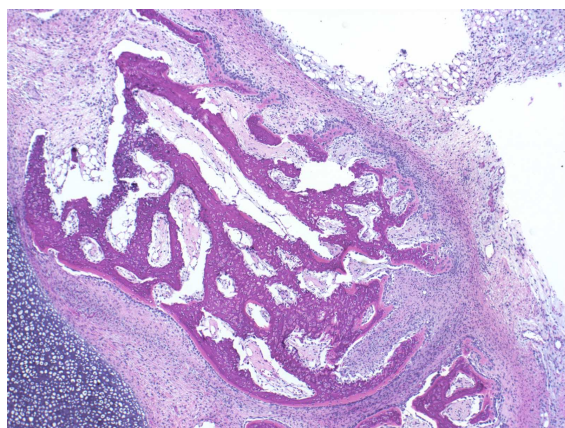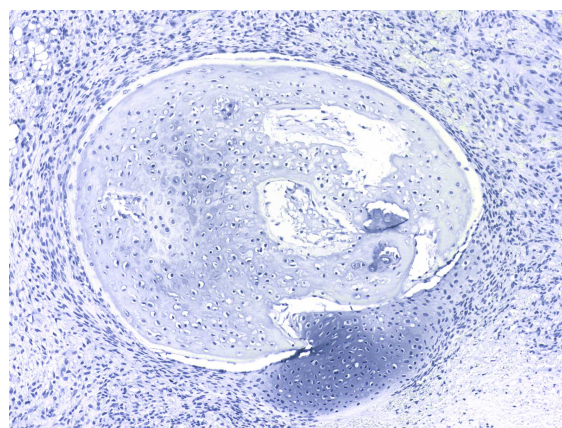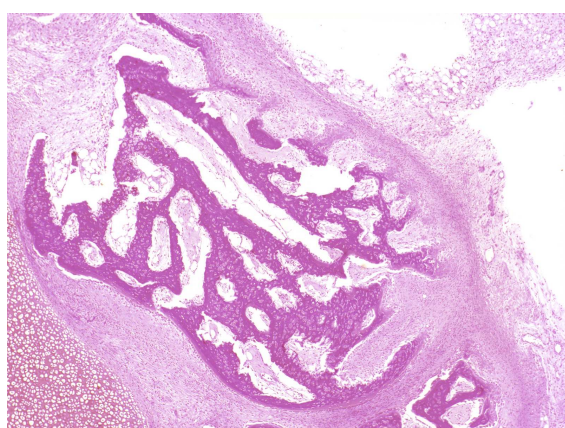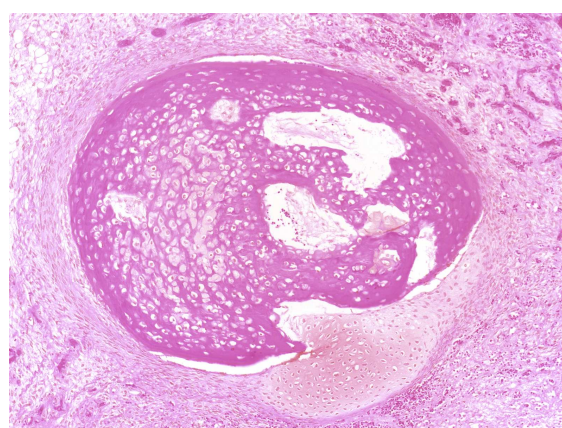
(g)
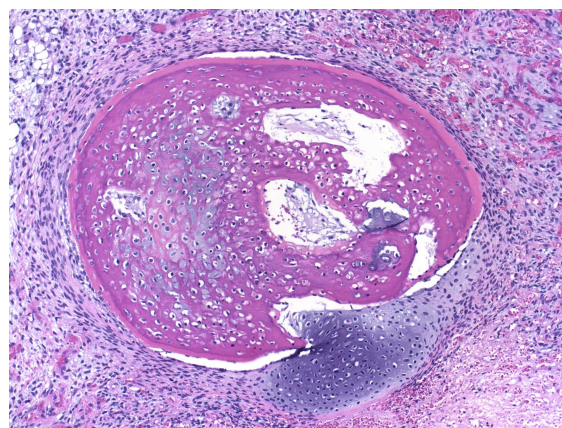
(h)
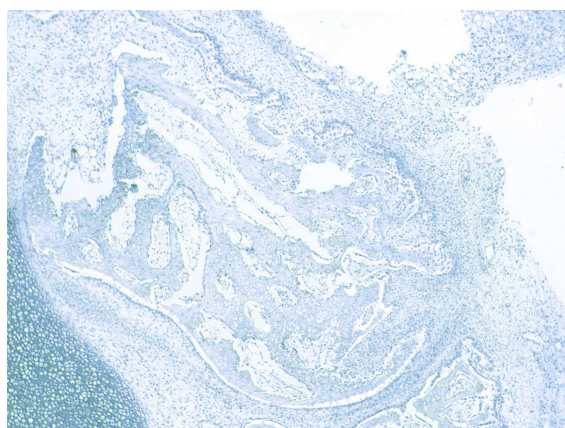
(i)

(j)

(k)

(l)

Figure 5.3 continued.

Chapter 6

Experimental Results

We now present the results of applying our overall classification algorithm to two very different histopathology applications of active colitis detection and teratoma tissue identification. In the first section we will discuss aspects of the experimental setup that are used for both applications such as feature selection, classifier design, training/testing set selection, and parameter choices. The next two sections present the experimental results for our applications with corresponding discussion. The last section focuses on overall experimental conclusions and discussion of various challenges and concerns we encountered during experiments.

6.1 Features

While the HV features are designed to work across applications, in each particular application only a subset of these features may be considered relevant. While a robust classifier may be able to automatically learn which subsets are appropriate for any given application, we choose to further leverage our HV methodology and select subsets based on pathologist's guidance. In future work, we hope to eventually incorporate automated feature selection into our algorithm.

Part of the definition of the HV features is the notion of the local neighborhood or, in our

case, the filter whose spatial span along with its values completely specify the neighborhood. As discussed before in Chapters 1 and 2, a tissue's or pathology's local presentation is the most consistent or stable in a particular range of sizes or scales. To learn about and capture this stable behavior, we explored families of filters whose members vary according to scale.

We investigated many standard image processing families such as simple averaging filters, edge detection filters (Prewitt, Canny, and Sobel), Gaussians and their derivatives, and Gabor filters. Averaging and standard Gaussian filters define easily understood neighborhoods; their shape and values result in a simple weighted sum of the feature (e.g., color or density) to be computed. The other families are less intuitive though their shape and values provide possibly useful qualities such as preferential orientation and frequency analysis. We found that due to a lack of knowledge about sample position and orientation, filters that have directional bias are ill-suited to the task. This observation combined with experimental results lead us to choose circular averaging filters, which are completely parameterized by their radius and the fact they sum to 1, and whose isotropy removes the effect of sample orientation.

Using this family of filters, we follow a scale-space type of approach [72] in attempting to learn ideal scale selection for a given application. Given a fine-resolution or very small scale, subsequent scales are derived by multiplying by powers of a fixed factor. Our experiments with low-magnification imagery revealed 2 to be a reasonable factor, with a maximum fine-resolution scale of 4 pixels. The final set of scales with which we experimented are $s = 2^k$, $k = 2, \ldots, 6$ pixels. This family of filters is used for all our subsequent experiments with references to scale being the radii of the circular averaging filter.

6.2 Classifiers

While the main focus in our work has been the intelligent design of features and a mathematical framework to support them, the overall goal is a maximally effective overall algorithm.

We have thus performed extensive experimentation in choosing a classifier among the typical ones such as SVMs, NNs, nearest-neighbor classifiers [44], and LDA methods. Our experiments have shown NNs to be the most effective for our applications due to their overall effectiveness and efficiency. SVMs also showed promise though their learning rate was far from ideal and often non-converging due to the sheer amount of data being addressed.

To quickly review NN basics, the algorithm is a series of layers which consist of weighted linear sums, constant biases, and function evaluations. There are three types of layers: input, hidden, and output. Each layer consists of a series of nodes, each of which evaluates the same pre-defined function or activation functions (for that layer) using the weighted linear sum , biased by some constant, of the previous layer's node values. The input layers is evaluated using the initial values input to the NN, i.e., the feature vector to be classified. Hidden layers are named due to the fact that their inputs and outputs are not directly visible. The output layer's outputs are the NN's overall result which, depending on the manner of training, indicates the determined label in a variety of ways.

The primary design choices of the NN are the number hidden layers used, the number of nodes in each layer, and the functions to be evaluated at each layer. We refer readers to [46] and other related works for more information on methods for determining these parameters. Once these aspects of the NN are fixed, the weights and bias used to compute the function input for each layer are learned based on training data and labels. The most popular method of learning is backpropogation [73], an iterative optimization method that updates an initial set of weights and biases (e.g., randomly initialized) to minimize the overall error between NN output and the desired training labels. The combination of number of layers, size of each layer, weights, biases, and functions allow a NN to learn a very complex function of the feature space that can be very effective. However, said complexity may also lead to overfitting to the data and increased sensitivity to variations in data.

As our prototype NN design, we use a simple 2-layer setup consisting only of an input

layer and an output layer. The input layer consists of as many nodes as the length of the feature vectors and hyperbolic tangent sigmoid activation functions. The output layer consists of as many nodes as the number of labels of interest with linear activation functions resulting in an output label vector as opposed to a scalar label. We use a 1-hot training approach where the label for each training sample is a binary vector where only the position for the appropriate label is non-zero. The final label is taken as that of the ideal 1-hot binary vector whose Euclidean distance to the output label vector is smallest. Our choice of network design is motivated by our past success with it in a variety of applications including histopathology and its overall simplicity which aids efficiency and helps avoid overfitting.

6.3   Training and Testing Set Selection

Training and testing set selection in our approach is a two-step process that considers the two functional units of data available, images and pixels. While the core unit of data is a pixel, one must consider the issues of selecting pixels without considering their source images. Our goal is to learn the identity of a tissue or pathology or other label of interest in a general manner so as to ensure that new examples are classified correctly. Consider the situation where all training pixels for a particular tissue come from a small set of images and furthermore, all available pixels in these images are used. In this situation, there is the chance that the overall algorithm will learn to identify these particular images as opposed to the tissue at large. In other words, overfitting may occur. To avoid this while maintaining a clear separation between training and testing data, we perform the following two-step process.

As is customary, the first step involves the separation of the available images into training and testing sets. We begin by randomly choosing (according to a uniform distribution) a fixed percentage of the available images that contain any of the labels of interest for training with the unselected images forming the testing set. From each training image, we randomly

sample a percentage of the pixels from each label of interest present. The resulting collection of sampled pixels across the training images for each label of interest form that label's training set. While the exact method of pixel sampling can vary, we have chosen a weighted random sampling that gives greater likelihood to pixels located in the interior of a a given label's region in any given image. This method is based on our observation that pixels at the boundaries between labeled regions, provide less reliable information that may confuse the classifier.

For a given application, due to the biological nature of the data, certain labels are likely to occur more often than others. For certain applications there may be no way to reliably collect more examples of a deficient label, making the effect of these imbalances not negligible. To address this issue, we enforce a level of equality among the training sets for each label by further sampling the training sets to the same size as the smallest training set formed so far. To avoid situations where this additional sampling results in overly small training sets for all labels, we also enforce a minimum training set size if it can be met by the available data.

Due to the sheer number of permutations of images, pixels, and associated training percentages, we have only explored a small set of training and testing set configurations. For the experiments presented, any given training and testing set pair are created by first sampling 50% of the available images for training with the rest assigned to testing. From each of the training images, 1% of the available pixels are sampled to create the training set. The minimum training set size was chosen to be 20 thousand samples to allow for maximum computational efficiency. No data from any of the training images is used for testing, maintaining a complete separation between training and testing data.

6.4 Parameter Selection

We now cover parameter choices made in our experiments; some of which we have discussed in Section 6.1, such as the choice of neighborhood via filter design; the classifier choice

we discussed in Section 6.2. Local histograms form the basis of the majority of our HV features, and in spite of often avoiding their explicit computation, properties such as their coarseness/fineness must still be defined.

For example, in quantifying local average color, we must choose a quantization which is equivalent to setting the number of bins in the local histogram given a defined range of colors. Sometimes the choice is implicitly made by the feature such as density where the local histogram only contains 2 bins corresponding to nuclei and non-nuclei pixels. When we do have the option of defining this aspect of the local histogram, we must consider the computational requirements discussed in Chapter 4. Specifically, given an image's spatial resolution and the available memory, how fine can our local histograms be without significantly compromising computational efficiency?

As mentioned earlier, trade offs can be made to alleviate these issues at the cost of information. As a result, parameter selection for local histogram formation is done in a two-step manner. First, given the computational constraints of the executing system, cross-validation within a number of training sets is performed to select optimal parameters. Given the large number of possible choices for these parameters, a relatively small set of variation is analyzed. For example, the number of quantization levels used in any given local histogram is usually chosen to maximize the usage of all available memory while satisfying the desired spatial resolution. The second step is a mass repetition of these experiments to learn the overall performance trend with respect to these parameters. Within the scope of this work, parameter selection is left to the cross-validation stage although as the work matures, this selection will be based on our increasing number of observations.

The other major parameter to be selected is the exact spatial support of scale of the neighborhood used to compute the local features. In spite of our guidance by pathologists, there is little intuition about what the ideal scales for any given tissue or pathology is. When combined with the intricacies of the algorithm, the scales at which classification can

be performed the most accurately and robustly for a given label can only be determined through empirical study. The experimental results we will present include such a study by observing the performance trends with respect to changes in scale. We not only attempt to identify ideal scales for individual tissues and pathologies, but also for groups of them where there are clearly differences in performance amongst individual labels.

6.5   Clinical Application: Active Colitis

To illustrate our algorithm on a clinical application, we attempt to identify an inflammation in the GI tract, more specifically of the colon, called *active colitis*. As detailed in Section 2.1.2, it is typically diagnosed by performing an endoscopy to obtain a biopsy of the inner wall of the colon. After sectioning and H&E staining, the biopsied sample is visualized with a microscope; the pathologist makes a diagnosis. While active colitis has many degrees of severity, we consider only a 2-class problem, that is, discrimination between the absence or presence of active colitis.

6.5.1   Dataset

The dataset consists of 40 colon biopsies that have been H&E stained and imaged at a resolution of 1600×1200 under two magnifications, 40X and 100X (see Figures 2.6 and 2.7). Half of the samples are considered normal (absence of inflammation) while the other half are active colitis (presence of inflammation) resulting in a total of 40 images for each magnification.

Each image is labeled as either normal or active colitis by our pathologists to provide a ground truth. A normal label is given only if the entire image is free of active colitis while an active colitis label is assigned if any active colitis is observed. This means that for a given biopsy, the labels of its two images may differ, but thanks to careful selection of the biopsy regions imaged, this is not the case with our data.

However, there is still some degree of uncertainty in the ground truth of the images labeled as active colitis. Specifically, active colitis images do demonstrate differing degrees of inflammation that translate to how much of the image is inflamed, e.g., an image where only 10% of its area is inflamed is labeled the same as an image where 90% of the area is inflamed. This lack of distinction in the ground truth will be discussed later when analyzing our results. Using our component segmentation, we detect no-stain regions in every image and assign them a background label (ignored for all purposes), with the rest of the image being given the pathologist's label.

6.5.2 Experimental Setup

In these experiments, we used the most appropriate subset of HV features based on expert guidance: nuclei density, ratio of lumen, and color of background/fiber colors. Figure 2.5 illustrates the differences between the two pathologies and their more discriminative qualities that coincide with the features selected. Testing of the HV feature sets at different scales to determine ideal spatial support, the classifier used, as well as training and testing set formation are as specified at the beginning of the section.

Though each image is given only a single label, we still perform pixel-level classification to avoid modification of our algorithm. Thus, the ideal result for any given image is one where every pixel is given the same and correct label; implications of this approach with respect to the nuances of the ground truth will be discussed later. We create 10 different training/testing sets (folds) using the previously discussed approach and report average pixel-level classification accuracies for each pathology of interest.

6.5.3 Results and Discussions

Table 6.1 gives average classification accuracy across magnifications and scales. The first trend we notice is that the performance on this dataset is relatively insensitive to the scale

of the feature set. We believe that this is because at these magnifications the cellularity, i.e., the nuclei of a region is still sufficiently captured. However, at higher magnifications, scale will be more important as too small an area will only capture a few nuclei and will not reveal the true identity of the sample.

Accuracy [%]	40X		100X	
Scale [pixels]	Normal	Active colitis	Normal	Active colitis
4	94	94	81	76
8	95	93	87	78
16	94	92	87	78
32	92	90	84	77
64	93	90	88	81

Table 6.1: Average pixel-level classification accuracies for normal and active colitis images over 10 distinct folds.

The next trend is that the accuracy is higher in identifying normal tissue; this is primarily a reflection of the feature set and the fact that our ground truth assigns one label to an entire image. For samples labeled as normal, the entire sample must be free of active colitis and thus the ground truth is more accurate; on the other hand, for samples labeled as active colitis, the pathology may not be present throughout, and thus the ground truth may not be completely accurate in a spatial sense. Given this lack of detail in the ground truth, our pixel-level classification approach can result in apparent misclassifications where regions labeled as being active colitis may in fact be normal. However, this same consideration must be given to the training of the algorithm where training images may be presenting normal tissue as active colitis. The overall result of this aspect of the ground truth on both training and testing has not yet been explored due to the lack of refinement in the ground truth.

Sill, in this application where a noticeable presence of active colitis is a call for treatment such an approach is not without merit. Our algorithm can identify all samples where more than 25% of the sample's area contains active colitis which correspond to all samples labeled

as having active colitis by our collaborating pathologists. We show in Figures 6.1 and 6.2 some examples of identification and delineation for each magnification and each pathology.

(a) Example 1 - Normal labeled 40X image

(b) Example 1 - Result

(c) Example 2 - Active colitis labeled 40X image
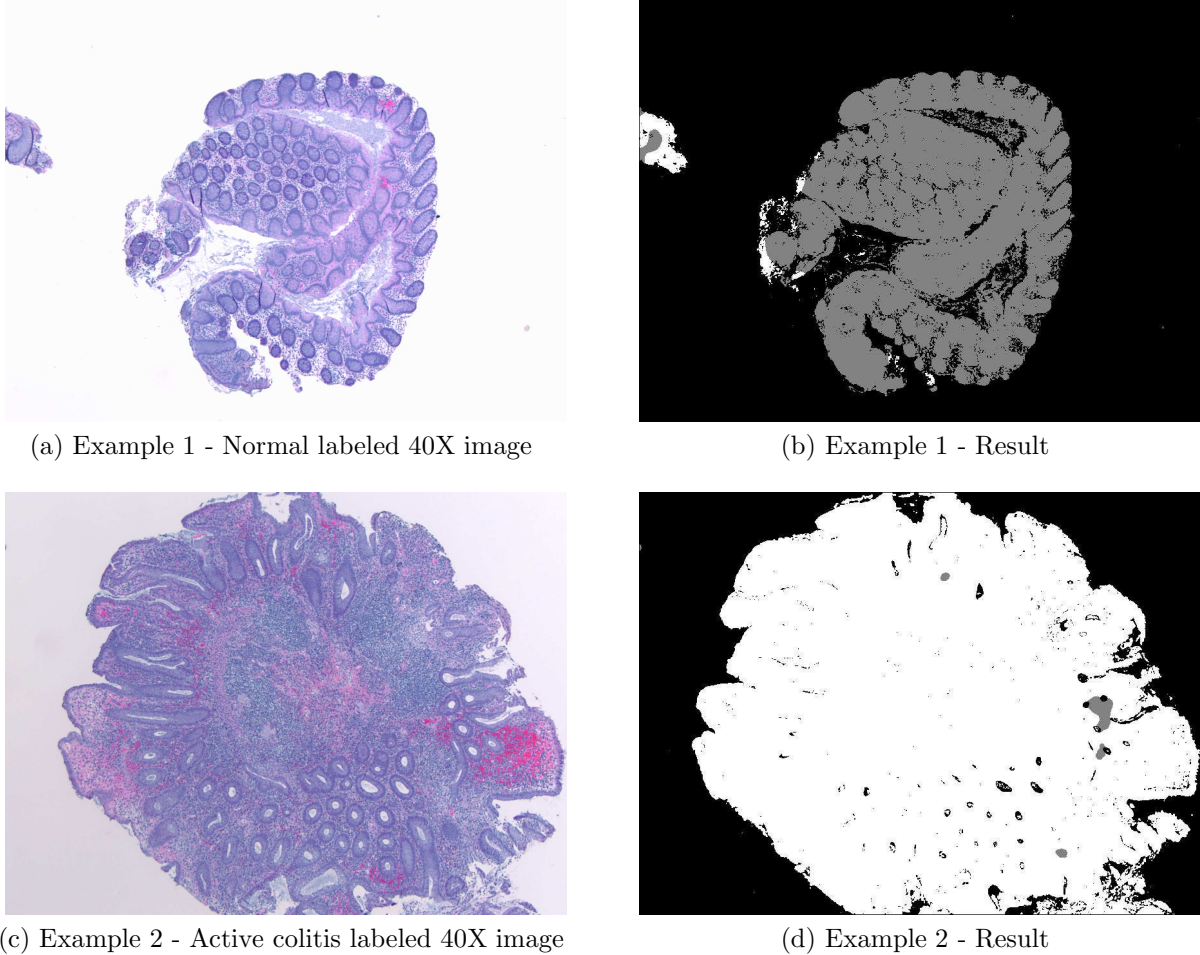

(d) Example 2 - Result

Figure 6.1: Example identification and delineation results for 40X magnification colon data. Black is background, gray is normal, and white is active colitis (ideally, normal should be all gray and active colitis should be all white).

While the results are promising, there are still aspects that need to be addressed, such as the introduction/implementation of architectural features describing various macro-structures present in a given sample. For example, the current feature set cannot quantify macro-structures such as crypts adequately (see Figures 2.2, 2.4, and 2.5). The first step in such a feature would be the segmentation of crypts which the work of [74, 75] addresses in the context of colon-related applications.
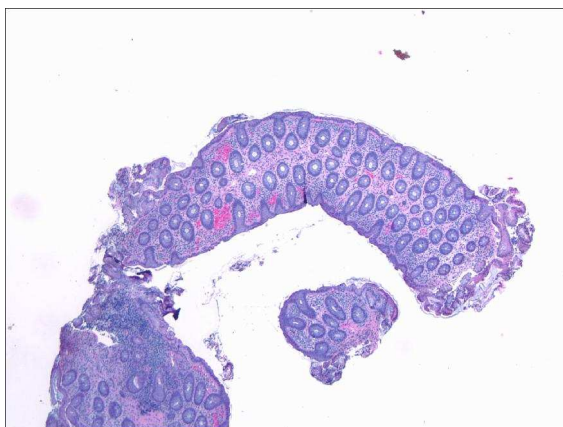
(e) Example 3 - Normal labeled 40X image
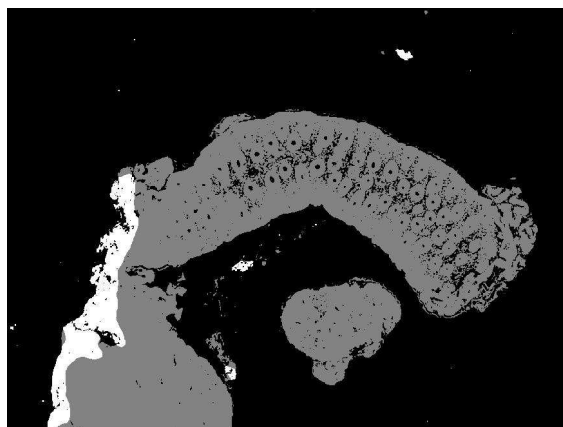
(f) Example 3 - Result
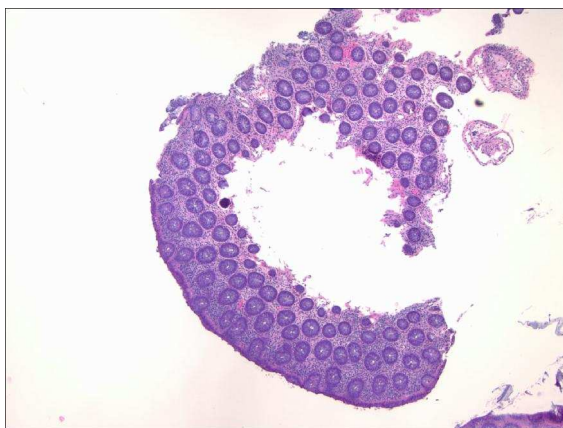
(g) Example 4 - Active colitis labeled 40X image
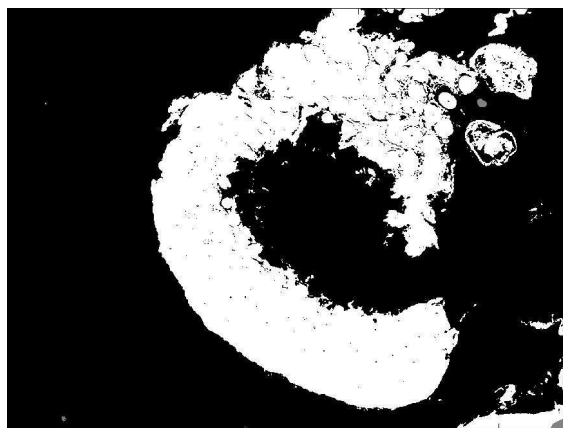
(h) Example 4 - Result

Figure 6.1 continued.

While ideally we would prefer a characterization of macro-structures that was independent of the application, this may be a less than optimal approach as a whole. In particular, the various sources of variation (acquisition and imaging specifics, sample to sample variance, etc.) may be better addressed by application specific methods for quantization of known macro-structures. While this somewhat contradicts our goal of creating a vocabulary to describe the large number of problems in this domain, it is still motivated by expert knowledge in that the particular macro-structures we wish to characterize are ones identified by collaborating experts. We expect the addition of such a feature to improve the accuracy in identifying active colitis; this is left for future work.

(a) Example 1 - Normal labeled 100X image

(b) Example 1 - Result

(c) Example 2 - Active colitis labeled 100X image

(d) Example 2 - Result
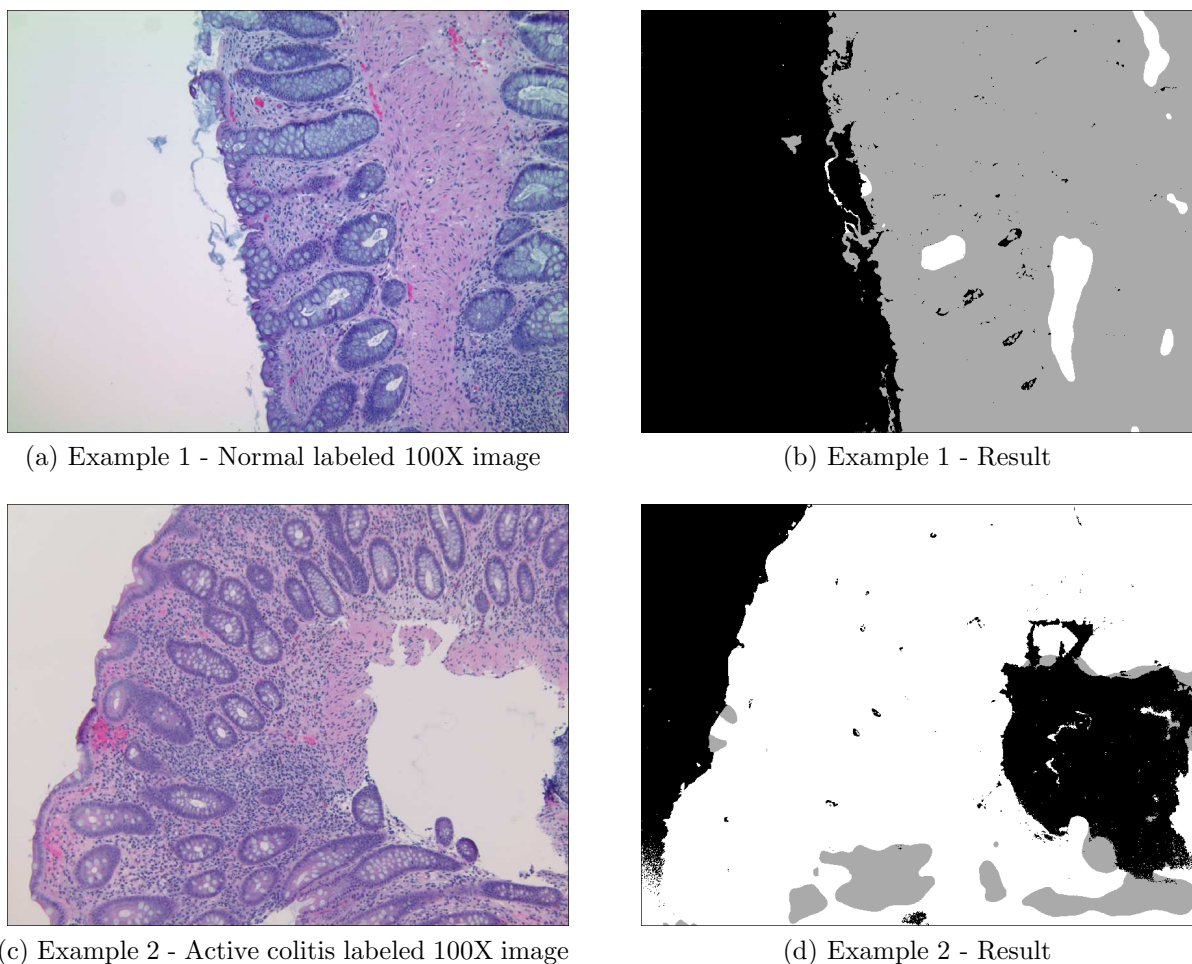
Figure 6.2:   Example identification and delineation results for 100X magnification colon data. Black is background, gray is normal, and white is active colitis (ideally, normal should be all gray and active colitis should be all white).

6.6   Research Application: Teratomas

6.6.1   Dataset

The dataset consists of H&E-stained samples of teratomas whose purpose is to research the development of germ-layer components from human and nonhuman embryonic stem cells (see Section 2.1.1). All tissue types found in humans can be found in this data, and individual images often consist of multiple tissue types and pathologies.(e.g., levels of maturity), making it one of the most general classification tasks possible in histopathology.
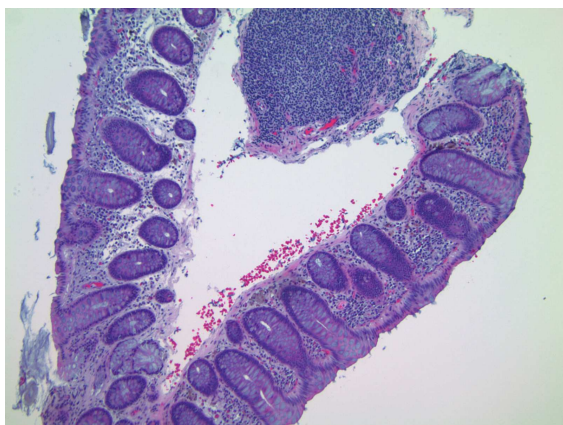
(e) Example 3 - Normal labeled 100X image
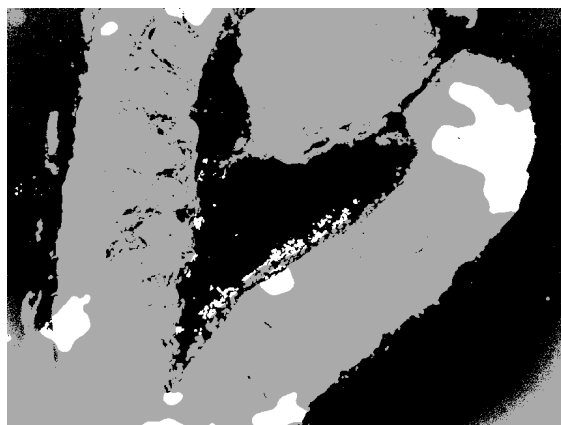
(f) Example 3 - Result
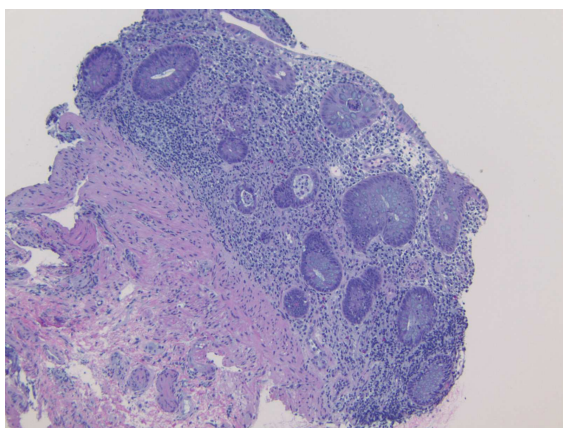
(g) Example 4 - Active colitis labeled 100X image
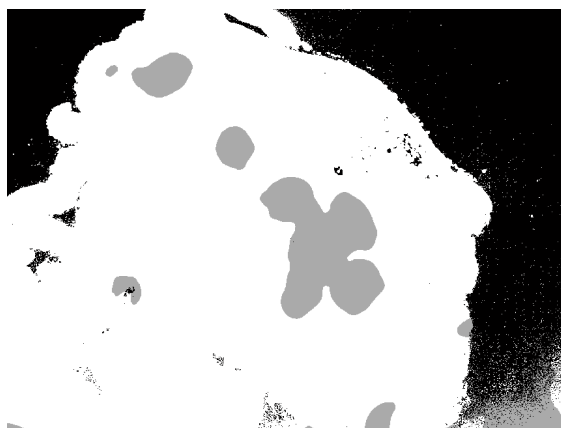
(h) Example 3 - Result

Figure 6.2 continued.

Teratomas are derived and serially sectioned as detailed in [49], then H&E-stained and imaged at 4X magnification resulting in 36 images of size 1600×1200; 15 tissues appear in these images although the number of images in which each tissue appears varies greatly. We choose to work with this magnification as it is the most useful from the pathologist's point of view as well as because it allows for the greatest amount of multiple tissue presentation. The ground-truth for each image is obtained by hand segmentation and labeling of tissues by our collaborating pathologists. Regions of uncertain identity, due to either lack of information or artifacts, are ignored and not used for evaluation.

6.6.2 Experimental Setup

Given the nature of this dataset, we have an opportunity to gauge the discriminative and generalization power of our algorithm and HV features. To this end, we perform a series of experiments in which the number of tissues being classified changes. We begin with the 2-class problem of classifying bone (B) and cartilage (C), chosen as our baseline due to the relative ease of distinguishing them. We then introduce immature neuroglial tissue (I) for a 3-class problem, neuroepithelial tissue (N) for the 4-class problem, and finally, fat (F) for the 5-class problem. Testing of the HV feature sets at different scales to determine ideal spatial support, the classifier used, as well as training and testing set formation are as specified at the beginning of the section.

Given the multiple-tissue composition of the images, this experiment represents a true test of our identification and delineation goal. We create 10 different folds using the previously discussed approach and report average pixel-level classification accuracies for each tissue of interest.

6.6.3 Results and Discussion

Table 6.2 gives average classification accuracy for each problem size and the best-performing scale, while Figures 6.3 through 6.7 shows some examples of identification and delineation.

Scale [pixels]	Accuracy [%]				
	B	C	I	N	F
16	89	89			
16	84	84	86		
32	78	81	75	71	
32	75	76	71	70	71

Table 6.2: Average pixel-level classification accuracies for teratoma images over 10 distinct folds with best-performing scale indicated for each problem size.
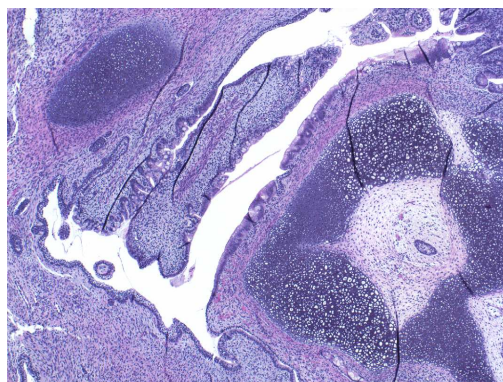
(a) Example 1 - Image
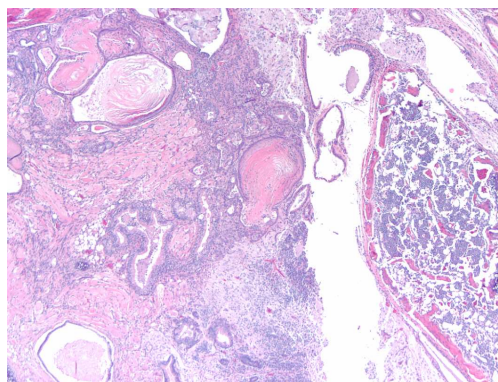
(b) Example 2 - Image
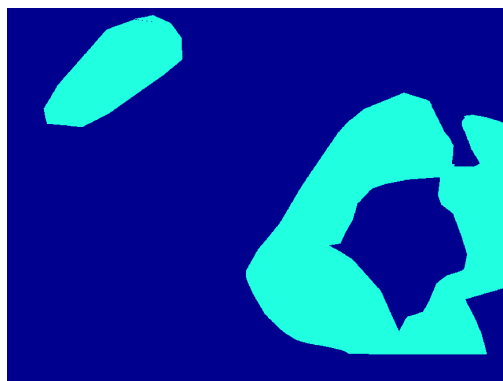
(c) Example 1 - Ground truth

(d) Example 2 - Ground truth
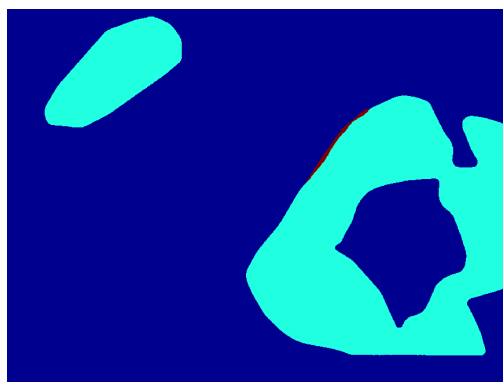
(e) Example 1 - Result
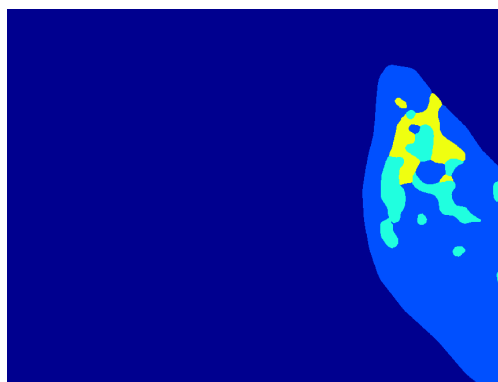
(f) Example 2 - Result

Figure 6.3: Example identification and delineation results for 4X magnification teratoma data. The ground truth and results are color coded as follows: background (dark blue), bone (light blue), cartilage (cyan), fat (yellow), neuroepithelial (orange), and immature neuroglial tissue (maroon).
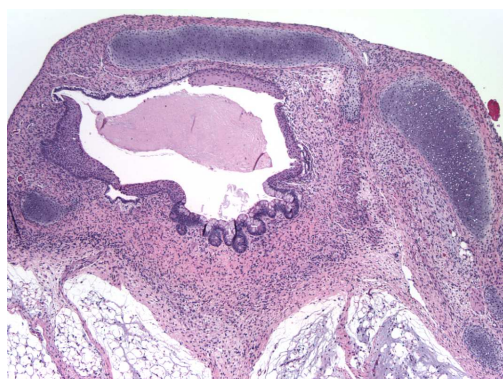
(a) Example 3 - Image
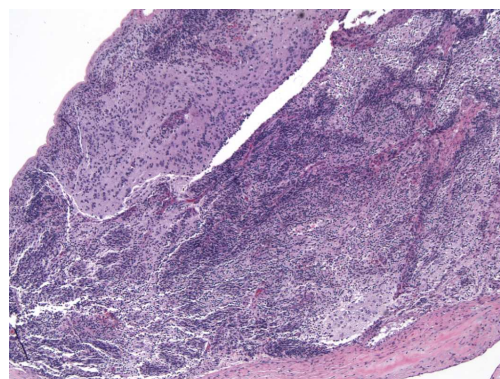
(b) Example 4 - Image
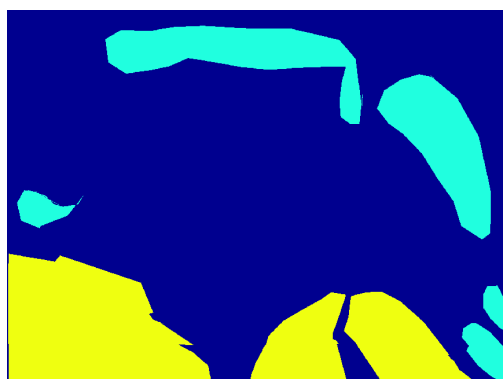
(c) Example 3 - Ground truth
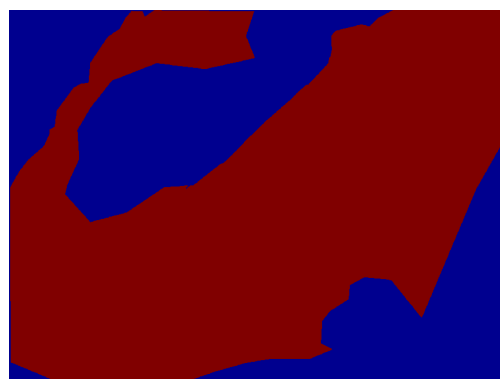
(d) Example 4 - Ground truth
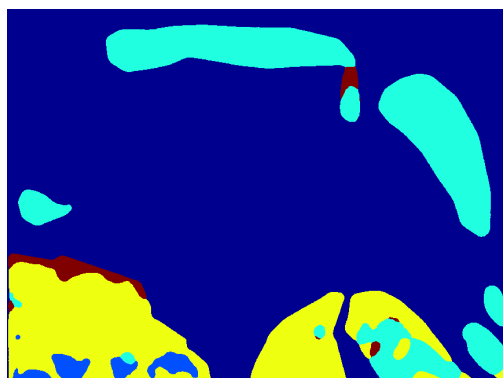
(e) Example 3 - Result
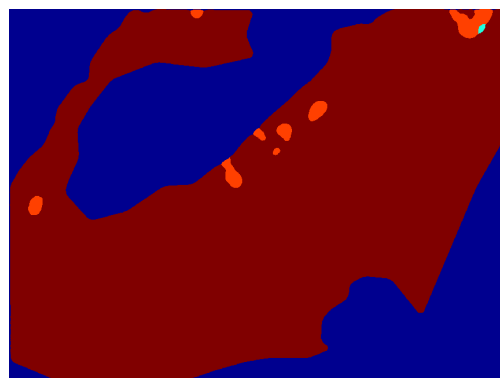
(f) Example 4 - Result

Figure 6.4: Additional examples of identification and delineation results for 4X magnification teratoma data.

(a) Example 5 - Image

(b) Example 6 - Image

(c) Example 5 - Ground truth

(d) Example 6 - Ground truth

(e) Example 5 - Result

(f) Example 6 - Result
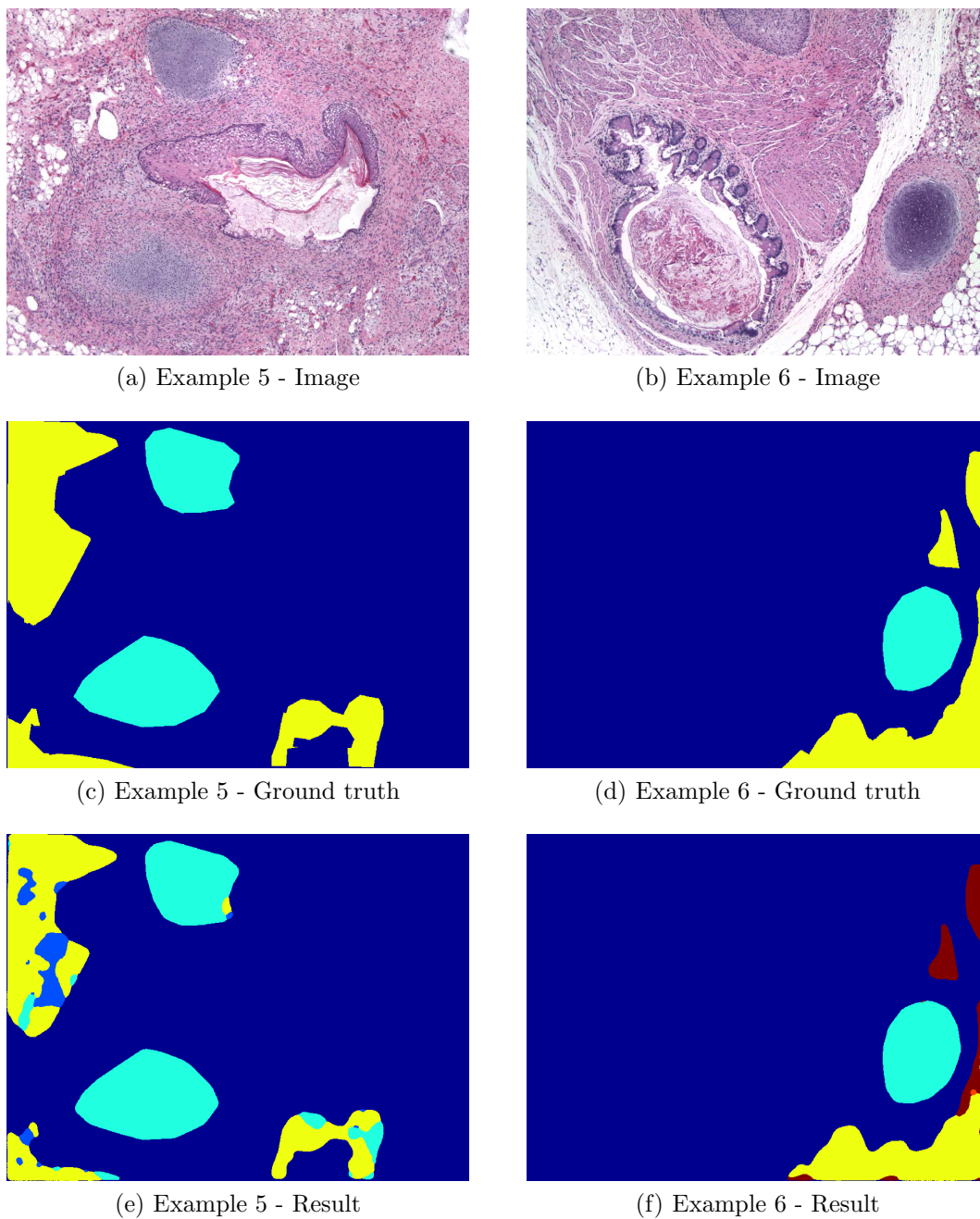
Figure 6.5: Additional examples of identification and delineation results for 4X magnification teratoma data.

(a) Example 7 - Image

(b) Example 8 - Image

(c) Example 7 - Ground truth

(d) Example 8 - Ground truth

(e) Example 7 - Result
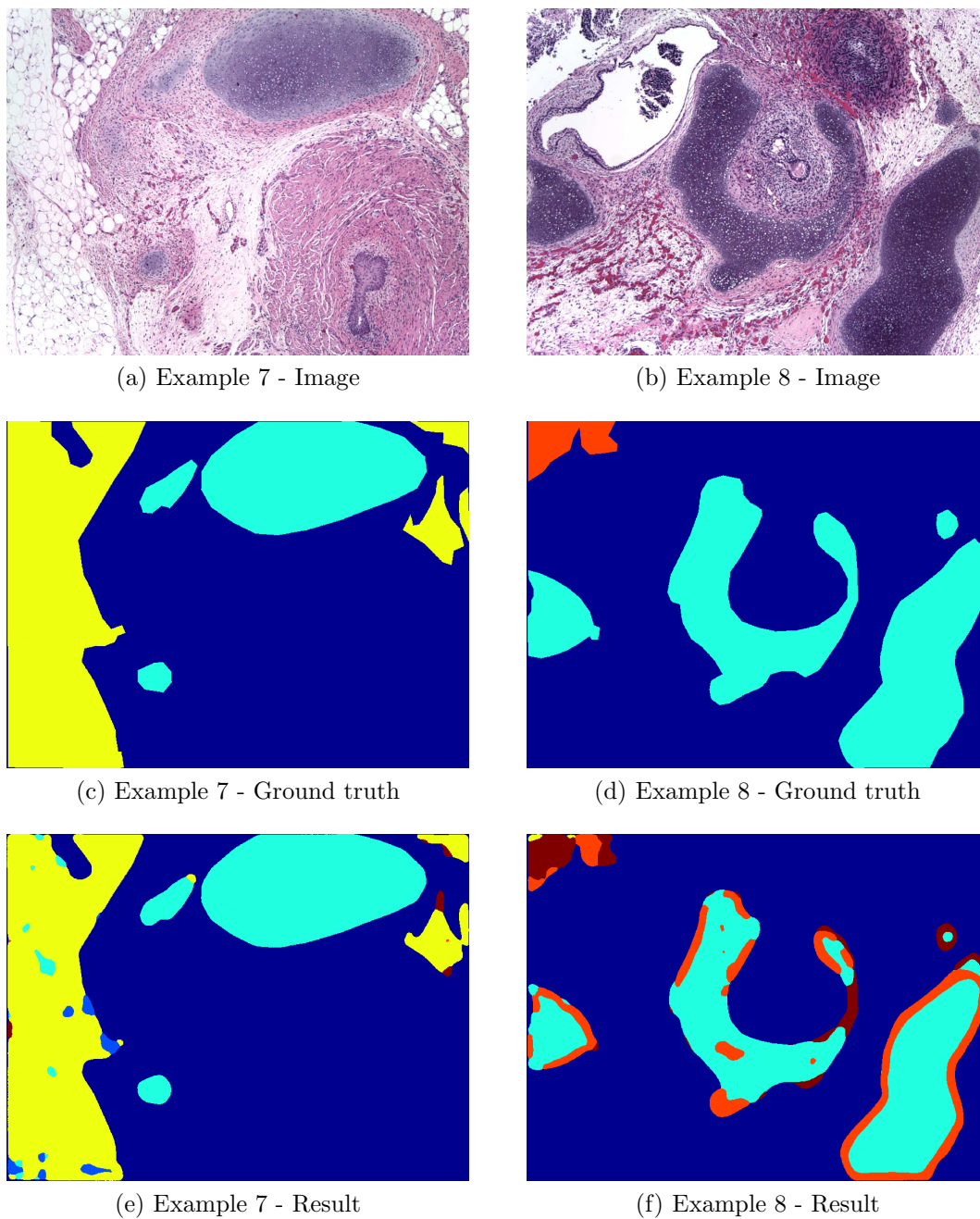
(f) Example 8 - Result

Figure 6.6: Additional examples of identification and delineation results for 4X magnification teratoma data.
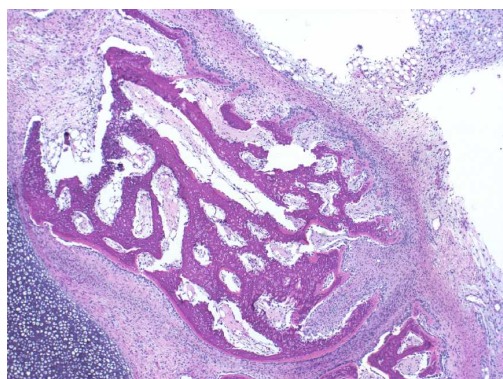
(a) Example 9 - Image
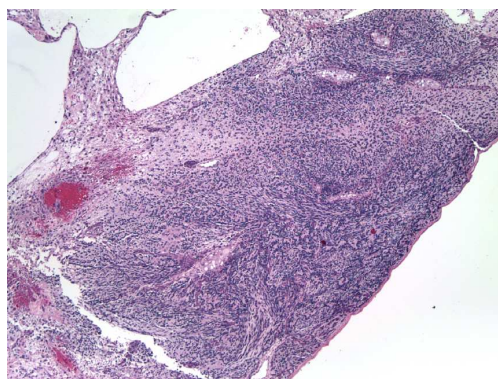
(b) Example 10 - Image
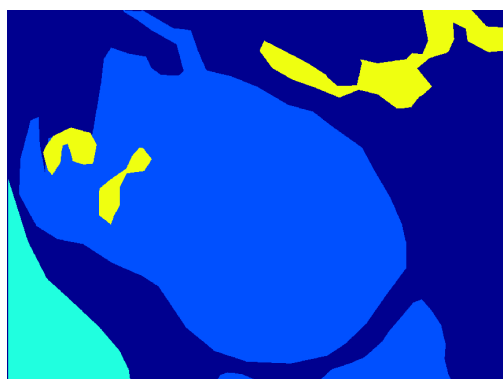
(c) Example 9 - Ground truth

(d) Example 10 - Ground truth
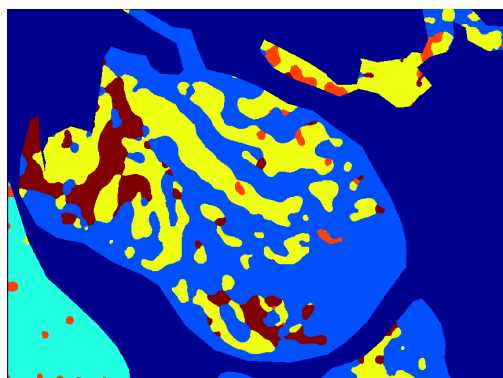
(e) Example 9 - Result
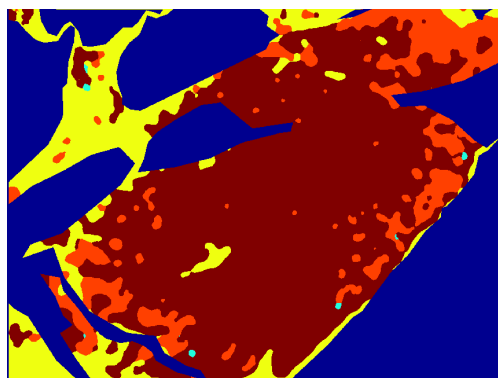
(f) Example 10 - Result

Figure 6.7: Additional examples of identification and delineation results for 4X magnification teratoma data.

The trend we observe is that as the problem size grows the performance degrades. This is not unexpected due to the difficulty of the problem. Based on various confusion matrices, bone and cartilage are consistently well separated with misclassifications coming from the other three classes. The majority of the HV features capture visual cues though some focus on architectural/structural features. This division of discriminative power is observed in the results. Tissues that are primarily distinguished based on visual cues such as cartilage and bone perform better than those where architectural cues are more important such as neuroepithelial tissue. Naturally, as more appropriate features are implemented, we expect performance amongst these tissues to improve and equalize.

The scale preference of various problems may demonstrate some consistency in the presentation of various tissues or the scale at which differences are maximized. These empirical results suggest some optimal scales for 4X magnification images. Specifically, it seems that the scales of 16 and 32 pixel radii are appropriate for this collection of tissues at 4X magnification. With respect to the other scales, these relatively medium scales indicate that the discriminative information is spatially localized. We are in the process of collecting images of samples at multiple magnifications to determine if scale preferences change proportionally to magnification.

The apparent fragmentation of some tissues highlights the need for architectural features in the HV. For example, as seen in Figure 6.7, bone is fragmented, primarily by labels of fat, mainly due to the lack of features quantifying the spatial relationship between lumen and bone tissue. In contrast, the delineation of cartilage is relatively smooth. This is due to cartilage's homogeneous appearance in a given image and the relative scale-invariance of its architectural cues.

Scale [px]	Accuracy (HV | Gabor | LBP | Textons) [%]				
	B	C	I	N	F
4	72 | 35 | 62 | 40	72 | 51 | 65 | 60	66 | 44 | 50 | 54	**73** | 33 | 60 | 39	65 | 44 | 62 | 51
8	72 | 38 | 61 | 41	71 | 48 | 66 | 52	71 | 43 | 52 | 48	72 | 37 | 61 | 43	67 | 40 | 64 | 45
16	73 | 36 | 65 | 41	73 | 50 | 65 | 53	70 | 46 | 56 | 50	70 | 35 | 62 | 40	69 | 40 | 62 | 42
32	**75** | 38 | 58 | 42	**76** | 53 | 60 | 56	**71** | 48 | 50 | 52	70 | 38 | 59 | 42	**71** | 42 | 60 | 45
64	70 | 39 | 57 | 45	74 | 55 | 60 | 58	69 | 50 | 51 | 54	71 | 40 | 54 | 45	70 | 43 | 55 | 43

Table 6.3: Comparison of average accuracies of HV versus Gabor, LBP and texton features on teratoma data over 10 folds across scales. For each scale and tissue, the HV features perform the best. The best-performing scale for each tissue is given in bold.

6.7 Comparison to Other Methods

We designed the HV to mimic those tissue attributes pathologists use when identifying a given tissue. While one can argue that most of those features are not new per se, the success of the HV in identifying H&E-stained tissues clearly shows it is the right approach. In many histopathology classification tasks [6, 76, 77], texture information was chosen as the primary descriptor of class identity, as it is both intuitively and empirically a vital descriptor in many applications. We now compare our HV to commonly used texture descriptors: Gabor filter banks, local binary patterns (LBP), and textons. In the interest of space, we only consider the 5-tissue teratoma experiment. The choice of scales, classifier, training and testing sets are all identical and as described previously.

6.7.1 Gabor Filters

Gabor filters have been used in image analysis for a long time [78–84] including in histopathology applications [6, 76, 85, 86]. They are designed to respond to textures at specific orientations and scales. We use Gabor filters at 8 evenly spaced orientations $\theta \in \{0, \pi/8, \ldots, 7\pi/8\}$ over 5 scales $s = 2^k$, $k = 2, \ldots, 6$ (in pixels), for a total of 40 filters. Each of these 40 filters is applied to each channel of the RGB version of the image. The final features are formed by

computing the local mean, standard deviation, and mode of each filter's response, yielding a total of $40 \times 3 \times 3 = 360$ Gabor features per pixel.

From Table 6.3 we see that the HV features (first column) outperform the Gabor ones (second column) across all scales and tissues. Note that the Gabor features perform the best on cartilage; this coincides with our general perception that texture is an important distinguisher of cartilage (see Figure 5.2(e)). Similarly, the relatively low performance with respect to bone tissue reflects a lack of color and architectural information that is more indicative of this tissue as opposed to others. We thus conclude that while Gabor features are a reasonable texture descriptor in this case, they are still not powerful enough for high performance in this challenging application.

6.7.2 Local Binary Patterns

LBPs are another powerful method for texture characterization; initially proposed as a means to discriminate between textures using statistical measures [87], LBP methods have been used in many applications [34,88–91], including biomedical [92–94] and specifically histopathological applications [77]. The basic premise of a LBP feature is to describe a micro-texture (local texture) using simple spatial operators and encoding methods [34,87]. The first parameter of an LBP operator is a spatial neighborhood around a pixel of interest. The neighborhood values are thresholded against the value of the pixel of interest resulting in a binary vector that is then typically encoded using a binomial scheme (identical to binary to decimal conversion) to yield a unique texture identifier. This encoding is further refined by introducing rotation invariance and a sense of uniformity. For our comparison, we use a LBP operator with a circular neighborhood of radius 1 pixel with 8 evenly spaced (angularly) points. The resulting binary vectors are then encoded as rotation-invariant uniform textures and then further aggregated into local histograms of these codes. These local histograms are computed over the same regions of support as the HV, specifically, circular neighborhoods of

varying radii. These local histograms are then used directly as feature vectors for training and testing of classifiers.

From Table 6.3 we see that LBP (third column) perform well and better than Gabor features. We again see that while texture description is vital, it is not sufficient. In particular, color information and organization of components are needed. Note that the LBP features perform better at smaller scales; this fits our intuition on the power of LBP in quantifying micro-textures, for example, cartilage.

6.7.3 Textons

The term texton was first proposed by Julesz [95] and later mathematically modeled by Leung and Malik [96] for texture classification who proposed the use of filter banks and clustering methods to create a prototype library of textures with which to both classify and generate textures. Beyond texture classification, textons were used in a host of applications [97–99] including biomedical ones [100–102]. Typically, texton formulation begins with the application of a filter bank to a set of training images. This set of responses is then used to identify the most discriminative subspaces of the filter space using a variety of methods such as clustering or dimensionality reduction. A specified number of subspaces, or textons, is identified to form a dictionary. Test images undergo the same filtering process; their responses are compared to the dictionary to determine the most similar texton. These similarities are most commonly used to form histograms of texton occurrence within a spatial domain. The histograms of labeled images are then compared to those of unlabeled images with the best match dictating the assigned label. We use an implementation consisting of isotropic Gaussians and Laplacians of Gaussians in addition to oriented Gabor filters as before, for a total of 50 filters. Following the filter phase, we apply a K-means clustering using a Euclidean distance metric to learn a fixed number (we chose 100) of textons/dictionary elements. We then compute local histograms of texton occurrence in circular neighborhoods

of varying radii. These local histograms become the feature vectors used for training and testing.

From Table 6.3 we see that textons (fourth column) perform somewhere between LBP and Gabor features, and are inferior to HV. Note that here, commonalities between the texton and Gabor features become apparent; the improved performance at larger scales can be attributed to an increased robustness to noise. Moreover, the superior performance of textons when compared to Gabor features lends credence to the notion that the increased conciseness of the texton representation improves discrimination.

6.8   Rejection Experiments

As discussed in Section 5.3, rejection is a needed component for identification and delineation in some histopathology applications such as our teratoma data. To achieve this rejection, we investigated a variety of techniques of two types of approaches: rejection based on perceived confidence and rejection via classification. In this section, we will briefly discuss these techniques, present some results, and discuss our findings. In all our experiments, we use the exact same experimental setup that led to Table 6.2 to allow for direct comparison.

Under the first approach, we experimented with confidence thresholding and ensemble voting. Confidence thresholding is merely the rejection of a label if the associated confidence does not meet a defined threshold. The intuition behind this approach is that the overall classification algorithm will overwhelmingly learn the defined set of labels of interest almost to the degree of over-fitting. This strong learning should then result in the classifier being very confident for the labels it has learned but much less confident for everything else. However, this does assume that the classification region that has been defined by the classifier contains only the labels of interest and that the data to be rejected lies reasonably far away from this region. Should either of these assumptions be violated by some sample, the assigned label may have a high enough confidence to be incorrectly accepted instead of rejected. Given

our NN classifier training approach, our confidences are the reciprocal absolute Euclidean distance of the ideal 1-hot binary vector to the output label vector. Thus, smaller distances to the ideal label result in higher confidences and vice versa for larger distances. An empirically optimal confidence threshold of 0.7 (labels whose confidences are less than this are rejected) led to the results presented in Table 6.4. Variations of the threshold demonstrated extreme sensitivity by leading to either complete rejection of all testing data or complete acceptance, implying that this approach given the current classifier is ineffective.

Scale [pixels]	Accuracy [%]					
	B	C	I	N	F	R
16	74	72				24
16	70	69	65			28
32	65	66	63	60		26
32	62	61	60	59	57	30

Table 6.4: Average pixel-level classification and rejection (R) accuracies for teratoma images using confidence thresholding over 10 distinct folds with chosen scale indicated for each problem size.

The other technique of ensemble voting is still a type of probabilistic approach but takes advantage of behavior observed in boosting [103] and bagging [69] methods. The basic idea is that a collection of classifiers, each trained differently, will each make an independent decision about any given sample from which a final collective decision will be made. The manner in which the collection of decisions is combined into a single one is a topic of much research, but in the interest of simplicity, we take a simple majority vote approach. A computationally feasible number of 20 NN classifiers are trained, each using a different sampling (with replacement) of the available training data. Each of the NNs makes a decision with the final decision being the final label that was in the true majority, i.e., at least half of the NNs agreed. Those samples for which no label received at least half of the votes are rejected as a form of confidence. This approach is equivalent to bagging though we do

not use all the subtleties of that method. Using the same experimental setup as before, we present the results using this approach in Table 6.5. These results are an improvement from those of confidence thresholding, but still demonstrate an inability to simultaneously achieve reasonable classification of the labels of interest while rejecting everything else.

Scale [pixels]	Accuracy [%]					
	B	C	I	N	F	R
16	77	75				34
16	72	70	66			32
32	68	65	63	63		35
32	65	63	60	61	60	40

Table 6.5: Average pixel-level classification and rejection (R) accuracies for teratoma images using ensemble voting over 10 distinct folds with chosen scale indicated for each problem size.

The second approach of rejection via classification is generally a much richer area of research though it is also often hindered by complex algorithms and computational issues. The fundamental notion behind these types of methods is that the classifier will not only characterize the various classification regions but also the rejection space in an balanced manner. We chose to focus on a modifications to the standard SVM that include the 1-class SVM, the support vector rejection decision machine (SVRDM), and the support vector data descriptor (SVDD). As the techniques that we will present are highly complex, we refer readers to the referenced works for more detail.

The 1-class SVM [104] is a simple extension of the standard 2-class SVM. Rather than try to maximize the margin between two differently labeled sets of data, the 1-class SVM attempts to fit a hypersphere around as much of the data of each class; thus, for a fixed number of labels, we have as many hyperspheres. Discrimination between multiple labels using 1-class SVMs involves testing the membership of a sample to each label by measuring its distance from the centers of the hyperspheres. The final labels is assigned by choosing

the hypersphere, and its associated label, for which the sample lies not only within the hypersphere's boundary but also is the closest amongst all such hyperspheres. Rejection is accomplished by rejecting any sample for which no hyperspheres are found to contain the sample. Due to its derivation, the SVM extension to a kernel-based derivation applies. For our experiments we used a series of standard 1-class SVMs of radial basis function kernels and slack parameters computed automatically based on the nature of the training data; the results are summarized in Table 6.6. As can be seen, almost no rejection is accomplished using this method which implies that the hyperspheres are not only encompassing the labels of interest but also those labels to be rejected and the unknown data. Furthermore, the classification accuracies also suffer implying that the hyperspheres are also overlapping with one another.

Scale [pixels]	Accuracy [%]					
	B	C	I	N	F	R
16	50	52				6
16	45	44	40			8
32	40	42	38	36		9
32	38	39	36	35	35	13

Table 6.6: Average pixel-level classification and rejection (R) accuracies for teratoma images using 1-class SVMs over 10 distinct folds with chosen scale indicated for each problem size.

The next technique is the SVRDM proposed by Yuan and Casasent [105]. The premise of the SVRDM is to adapt the standard SVM to define the traditional positive and negative regions of classification in addition to a new rejection region. This adaptation is achieved by modifying the optimization metric typically used in the SVM to define a single boundary to one that defines two boundaries. One side of one boundary corresponds to the positive classification region while the complementary side of the other boundary defines the negative classification region. The region between these boundaries is considered the rejection region as samples lying in this area are not well-defined members of the defined labels. To achieve

rejection for multiple labels using this method, we train a SVRDM for each label where the label is the positive class and all other labels are collectively the negative class. A testing sample is assigned the label of the SVRDM for which its membership in the positive region of that SVRDM is the strongest. Rejection of the sample occurs when the sample does not belong to the positive region of any of the SVRDMs. The results using this method are summarized in Table 6.7 where a similar trend with the 1-class SVM is observed though the classification accuracies have improved.

Scale [pixels]	Accuracy [%]					
	B	C	I	N	F	R
16	56	55				5
16	52	51	48			7
32	50	51	43	42		11
32	49	47	40	41	40	15

Table 6.7: Average pixel-level classification and rejection (R) accuracies for teratoma images using SVRDMs over 10 distinct folds with chosen scale indicated for each problem size.

The final method is that of the SVDD [106] whose purpose is very similar to the 1-class SVM though the optimization method taken differs. Rather than try to maximize the amount of data captured by the hypersphere, the SVDD attempts to minimize the radius of the hypersphere that encapsulates a specified proportion of a label's training data. The difference, though apparently subtle, can lead to significant differences in the hypersphere found in the SVDD when compared to the 1-class SVM. Due to their similar nature, we take the same approach of classification and rejection as we did with the 1-class SVM and present the results in Table 6.8. The SVDD performance is almost equivalent to the SVRDM though it does achieve slightly improved rejection while maintaining similar levels of classification performance.

From these experiments, the difficulty of the rejection task is clear for the teratoma data. Since the teratoma application is very difficult in any regard due to its highly chaotic nature,

Scale [pixels]	Accuracy [%]					
	B	C	I	N	F	R
16	55	57				10
16	53	52	46			15
32	49	51	45	43		18
32	45	48	42	40	42	20

Table 6.8: Average pixel-level classification and rejection (R) accuracies for teratoma images using SVDDs over 10 distinct folds with chosen scale indicated for each problem size.

we will not claim that rejection is not feasible in other applications. Instead, we emphasize the lack of clarity in both defined labels and in seemingly unknown tissues that is the result of the biological nature of the data. Many other rejection approaches exist and we intend to explore them incrementally, though based on our results, the optimal direction is not clear.

Chapter 7

Discussion, Conclusions, and Future Work

7.1   Discussion

The overall goal of this work was to create a system for automated delineation and identification of H&E-stained tissues for digital histopathology. One of the two main guiding principles was to design and use a vocabulary of features that would mimic those attributes pathologists use when making decisions. In other words, we were not aiming to design necessarily new features, but rather a set of features uniquely suited to this domain. The other guiding principle of pixel-level classification is not a novel concept, but is well suited to this domain and our task. In following both of these guiding principles, a mathematical framework based on local histograms and occlusion models for the analysis of local textures was formulated. Again, while not wholly novel concepts, their use as a mathematical framework for histopathology is. As a whole, this algorithm seeks to be maximally effective by proposing a rigorous mathematical foundation with which the knowledge of the pathologist can be verifiably modeled. Though we have presented an ongoing discussion of our motivations and solutions, we present here some additional points of interest.

7.1.1   Feature Dimensionality

We now discuss the effectiveness of various features with respect to their dimensionality. This analysis provides further evidence supporting our guiding principle that a concise and intelligently guided description can be far more effective than a bulky undirected collection of descriptors. Figure 7.1 shows a comparison of the results shown in Table 6.3 with an additional, combined set, consisting of Gabor, LBP and texton features.
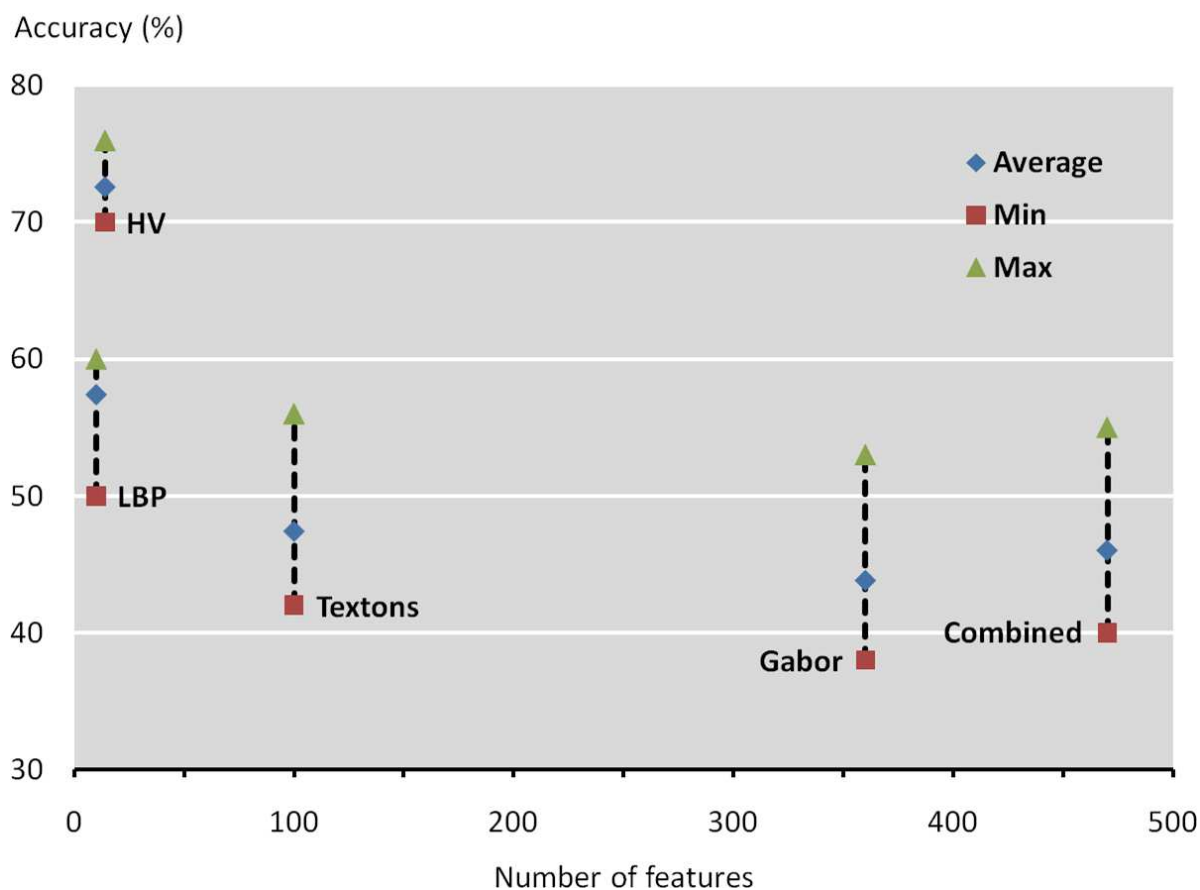
Figure 7.1: Comparison of results shown in Table 6.3 with the combined set consisting of Gabor, LBP and texton features. The blue squares represent the overall average accuracy across all 5 tissues while the red and green squares represent the minimum and maximum individual tissue accuracies respectively. All results are given for the best-performing scale of 32 pixels.

It is clear that the HV features outperform all other features while using a much smaller number of features (apart from LBP that is close in number). Furthermore, the simple combination of available features, as is often done in many machine learning applications, does not yield a tangible benefit. Based on the intervals of performance, one can infer that this combined feature set results in either an increased sensitivity of the overall algorithm or a case of overfitting.

As we discussed before, in spite of the variety of approaches, many algorithms do not achieve the same level of performance as their pathologist counterparts. In the pursuit of this performance standard, some researchers take the approach of combining large sets of relatively undirected features in the hopes of learning some combination of them best suited to the task at hand. While this approach may be valid in some situations (e.g., unknown classification problems), we believe it is not appropriate as a first attempt in the majority of histopathology applications. Rather, we must first learn and use the approach of the pathologist, i.e., those features that have been demonstrated to be effective and intuitive. Once this family of expert-guided features have been exhausted, we should then explore those undirected features.

It is at this limit of well-understood performance that we, as the designers of a computational algorithm, should utilize the exploratory power of machine learning methods for improvement. Assuming we have matched the pathologist's level of performance, it is these purely algorithmic efforts that will take us beyond human performance by exploring representations of the data that are not well-understood by humans. However, while we still pursue the pathologist's standard of performance in our applications, we believe that this family of problems can be best served by a sparse but intelligent and robust set of features.

7.1.2 Noise

A concern in any image classification task is noise, which in histopathology, occurs in a myriad of common and uncommon ways. Traditional image noise such as white noise, shot noise, and blurring can be attributed to the quality of the microscope optics and image sensors. Others sources of noise include variations in aspects of the physical acquisition such as the surgical procedure and the staining process whose effects are generally large physical defects and image-level variation, respectively. Since gross defects or image-level variations are not locally varying, our pixel-level classification approach will not be affected greatly by them. However, should these noise effects be so great that they truly alter or mask the identity of tissues and pathologies, a preferred approach is to identify such images as being of unacceptable quality.

However, noise that has a more local effect can have a negative effect on our algorithm. In particular, point-wise noise (e.g., white noise) and blurring (e.g., out-of-focus) would be serious challenges if the description of a pixel was based on the pixel alone. However, since our algorithm utilizes local descriptions for each pixel, we have some robustness to both these types of noise. While we have not explicitly quantified the noise present in our data (e.g., SNR), the relatively smooth (continuous) labellings we have achieved indicate that such image noise has not yet become a serious source of error.

7.2 Conclusions

Histopathology is a critical component of many clinical and research pathways in medicine. In particular, the immunohistochemical technique of H&E staining provides extremely detailed presentations of tissues and pathologies with which many important tasks can be performed. However, this same richness in detail means that only highly trained experts, i.e., pathologists, can properly perform these tasks. The constant need for histopatholog-

ical analysis results in a significant burden of time and effort on pathologists that can in turn lead to inefficiency, inconsistency, and inaccuracy in their analysis. To aid pathologists, engineers design computational algorithms that partially or completely address some of these histopathology tasks. This thesis proposes such an algorithm whose purpose is the automated identification and delineation of tissues and pathologies in H&E stained images, a task fundamental to many histopathology applications. The algorithm's theoretical and practical foundations not only help justify our approach, but also allow us to propose a scope much greater than those of other works in this area.

We began with our mathematical framework for describing certain types of local textures, such as the ones found in H&E stained images, for the purpose of classification. This framework defines local histograms as a unit of description and occlusion models as a method for generating increasingly more complex textures from simpler ones. We then derived relationships between the local histograms of complex textures and those of the simpler textures used to generate them via occlusion models. In addition to the theoretical contributions, the framework also provides methods for local histogram computation and occlusion models formulation. However, the potential power of local histograms is limited by their significant computational requirements, specifically, the large storage requirements. We concluded our mathematical framework by comparing it to the related works of mathematical morphology and the Dead Leaves Model where various similarities were shown, but fundamental differences in assumptions and applicability distinguish those works from our own.

Given our mathematical framework, we next presented algorithmic details beginning with a methodology for the creation of an expert-guided feature set that, for histopathology applications, we call the HV. The primary goal of the HV is to not only take advantage of the knowledge and experience of pathologists, but also improve the engineer's understanding of how the designed algorithm utilizes the features used. Without this understanding, we would be ill-equipped to understand the reasons for both positive and negative performance

from a feature perspective, and thus, find it harder to identify areas of possible improvement. Furthermore, these HV feature sets can be either specific or general in their applicability depending on what application they are derived from. For our experiments, we derive a HV features set for the very general application of tissue identification in teratomas and apply it to an additional application of active colitis detection.

The final portion of algorithm addressed how we achieve simultaneous identification and delineation through pixel-level classification. To avoid the exceedingly difficult, if not impossible, task of unsupervised segmentation of tissues and pathologies necessary for straightforward tissue identification, we classify each individual pixel based on it and its neighboring pixels. The use of a neighborhood or local features reflects the fact that a tissue or pathology is maximally stable in local appearance at certain scales. The previously defined HV features are formulated as local features to support out pixel-level classification approach. Through experimentation, we attempted to learn the ideal scales for a variety of tissues and pathologies.

We then applied our algorithm (utilizing our teratoma-based HV feature set) to two very different applications: detection of active colitis (a clinical task) and tissue identification in teratomas (a research task). Though these applications vary both in goal and scope, the same algorithm with minor modifications achieved more than reasonable performance for both of them. We compared our HV feature set's performance to those of some common, though general, feature sets used in other histopathology works and segmentation in general. In these comparisons, our more concise HV feature set significantly outperformed the other features, demonstrating the power of expert-guidance and the difficulty of this family of problems. Based on these results, we believe that our approach warrants further development to not only improve performance, but also to expand single-algorithm applicability.

However, while our final goal is the identification and delineation of labels of interest, we can only claim to have achieved promising success with identification while the delineation

task requires additional investigation. Specifically, given a particular application, portions of an image may need to be labeled as background or be rejected since they do not belong to the defined set of labels of interest. Thus, the algorithm needs the ability to reject these undesired samples, a task that is extremely difficult in any application. Our attempts at achieving this rejection met with varying types and degrees of success though no clear conclusion could be made. Fundamentally, the challenge of rejection in this application arises from a lack of data and the need to reject both structured and unstructured labels simultaneously.

This thesis recognizes and utilizes the simple fact that the general family of histopathology applications are based on tasks requiring highly specialized knowledge. To accurately and faithfully computationally model this knowledge requires both theoretical justification and verification of said computational modeling. Doing so maximizes our understanding of the algorithm that will be created, allowing us to not only better analyze its own performance, but also better identify the ways in which to improve upon it. Experimental results from different applications demonstrate promising performance though some aspects of our goal require additional research.

7.3 Future Work

7.3.1 Existing Algorithm

The work presented in this thesis represents the initial efforts in developing a highly capable and intuitive algorithm for identification and delineation of tissues and pathologies in H&E stained images. Our experimental results on two different applications encourages further research along many different directions including improvement of all aspect of the algorithm beginning with the mathematical framework.inning with the mathematical framework.

Our mathematical framework, though rigorous in its analysis, is lacking in some areas. First, methods for local histogram computation must be improved so that we can apply

our framework directly. Given the current state of computing hardware, improvements to software are the only reasonable path in the short-term. Storage is the primary limitation both as a strict constraint on the resolution of our local histograms and a computational hindrance. Compression of local histograms or implicit extraction of desired metrics from local histograms can reduce this storage constraint. In terms of computation, data and task parallelism are applicable given our definition of local histograms, allowing us to take advantage of distributed computing platforms. Though hardware will improve and relax these constraints in time, improvements in both computation and storage approaches can yield short-term gains.

We can also further our occlusion models by using them in a truly generative manner to create synthetic textures. The discriminative power of our framework has been shown through both theory and experimentation, but by demonstrating generative capability, we will "close the loop" between them. The challenge here is to not only model the simple textures that will represent basic components (e.g., nuclei and cytoplasm), but also determine the occlusion models that will accurately yield the complex textures of tissues and pathologies.

The HV methodology is simple enough that it does not require further improvement, but the features derived from its application can be improved. Specifically, our component segmentation algorithm should be made more robust to changes in appearance due to staining variations, exact magnification, and imaging conditions. Various HV features can be improved through the investigation of more appropriate and effective neighborhoods. New HV features can be extracted by applying the methodology to other applications, both clinical and research oriented.

We will continue to research rejection of unwanted labels in the context of H&E stained histopathology images. In particular, we plan to familiarize ourselves with the relevant literature in other fields such as automatic target recognition, biometrics, and image quality

assessment where rejection has long been needed and explored. The challenge of accomplishing rejection in a general and robust fashion lies in determining a rejection method that properly addresses the different natures of labels of non-interest and truly undefined labels.

7.3.2   Hierarchical Classification

Hierarchical classification is a new direction that could be quite powerful in not only improving the algorithm's performance, but also in expanding its applicability. As discussed in Chapter 1, sub-groups of the tissues and pathologies will demonstrate differing levels of similarity, one of the primary challenges of discriminating between the individual labels. When using a standard approach where all individual labels are identified simultaneously, members of a sub-group are often confused with each other because their perceived separation is small relative to the separation between the various sub-groups. To mitigate this type of error, we propose a divide-and-conquer approach where we incrementally divide a large group of labels into smaller and smaller sub-groups and update an assigned label in a similar fashion. The composition of these sub-groups should maximize the dissimilarity between them, and as a result, maximize the accuracy of classifying into these sub-groups. Each subsequent level of sub-group division corresponds to an increasingly fine classification, the final level being into the finest (smallest) sub-groups.

However, the choice of these sub-groups at any level is entirely dependent on the features used for classification at that level. For example, using color will lead to sub-groups that are similar in color. Our HV has provided an initial indication of the importance of features to the various labels and using this knowledge we could specify some intuitive division of labels. However, the empirical nature of the entire algorithm would require us to use some sort of feature selection in addition to our expert-guidance. After all these design choices are made, we would have a hierarchy where each subsequent level would refine the label assigned to a test sample using the optimal grouping of labels based on the optimal set of features for

that level. To determine this optimal configuration we could use an optimization approach though the challenges will be search space size, convergence issues, and expectations on final performance.

Appendix A

Proofs

A.1 Proof of Theorem 1

For (a), replacing x' with $-x'$, and substituting the relation $\delta_y(f(x - x')) = 1_{f^{-1}\{y\}}(x - x')$ into (4.1) yields:

$$
\begin{aligned}
(\mathrm{LH}_w f)(x, y) &= \sum_{x' \in \mathcal{X}} w(x')\delta_y(f(x + x')) \\
&= \sum_{x' \in \mathcal{X}} w(-x')1_{f^{-1}\{y\}}(x - x') \\
&= \sum_{x' \in \mathcal{X}} \tilde{w}(x')1_{f^{-1}\{y\}}(x - x') \\
&= (\tilde{w} * 1_{f^{-1}\{y\}})(x). \quad\quad\quad (\mathrm{A}.1)
\end{aligned}
$$

For (b), the definition of δ_0 gives:

$$
\begin{aligned}
[(\delta_0 \otimes \omega) * \mathrm{LH}_w f](x, y) &= \sum_{(x', y') \in \mathcal{X} \times \mathcal{Y}} (\delta_0 \otimes \omega)(x', y')(\mathrm{LH}_w f)(x - x', y - y') \\
&= \sum_{y' \in \mathcal{Y}} \omega(y')(\mathrm{LH}_w f)(x, y - y'). \quad\quad\quad (\mathrm{A}.2)
\end{aligned}
$$

171

Substituting (A.1) into (A.2) and using (4.4), gives our result:

$$
\begin{aligned}
[(\delta_0 \otimes \omega) * \mathrm{LH}_w f](x,y) &= \sum_{y' \in \mathcal{Y}} \omega(y')(\tilde{w} * 1_{f^{-1}\{y-y'\}})(x) \\
&= \sum_{y' \in \mathcal{Y}} \omega(y') \sum_{x' \in \mathcal{X}} \tilde{w}(x') 1_{f^{-1}\{y-y'\}}(x - x') \\
&= \sum_{(x',y') \in \mathcal{X} \times \mathcal{Y}} (\tilde{w} \otimes \omega)(x', y') 1_f(x - x', y - y') \\
&= [(\tilde{w} \otimes \omega) * 1_f](x,y).
\end{aligned}
$$

A.2 Proof of Proposition 2

For any $w \in \ell(\mathcal{X}, \mathcal{R})$ and $f \in \ell(\mathcal{X}, \mathcal{Y})$:

(a) Let $(x,y) \in \mathcal{X} \times \mathcal{Y}$. By exchanging sums and noting that for any fixed $x' \in \mathcal{X}$, $f(x+x')$ is equal to one and only one $y \in \mathcal{Y}$, it follows that

$$
\begin{aligned}
\sum_{y \in \mathcal{Y}} (\mathrm{LH}_w f)(x,y) &= \sum_{y \in \mathcal{Y}} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f(x + x')) \\
&= \sum_{x' \in \mathcal{X}} w(x') \\
&= 1.
\end{aligned}
$$

(b) Let $(x', y) \in \mathcal{X} \times \mathcal{Y}$. By definition of the local histogram,

$$
\begin{aligned}
(\mathrm{LH}_w \mathrm{T}^x f)(x', y) &= \sum_{x'' \in \mathcal{X}} w(x'') \delta_y(\mathrm{T}^x f(x' + x'')) \\
&= (\mathrm{LH}_w f)(x' - x, y) \\
&= (\mathrm{T}^{(x,0)} \mathrm{LH}_w f)(x', y),
\end{aligned}
$$

obtaining the result.

(c) For any $(x, y') \in \mathcal{X} \times \mathcal{Y}$, noting that $\delta_{y'}((f + y)(x + x')) = \delta_{y'-y}(f(x + x'))$ yields:

$$[\mathrm{LH}_w(f + y)](x, y') = \sum_{x' \in \mathcal{X}} w(x')\delta_{y'-y}(f(x + x'))$$

$$= (\mathrm{LH}_w f)(x, y' - y)$$

$$= (T^{(0,y)}\mathrm{LH}_w f)(x, y').$$

(d) Let $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $q \in \ell(\mathcal{Y}, \mathcal{Y}')$. An exchange of sums and the observation that for any fixed $y' \in \mathcal{Y}'$, $\delta_{y'}(q(f(x + x'))) = \delta_y(f(x + x'))$ for all $y \in \mathcal{Y}$ such that $q(y) = y'$, allows us to show that:

$$[\mathrm{LH}_w(q \circ f)](x, y') = \sum_{x' \in \mathcal{X}} w(x')\delta_{y'}(q(f(x + x')))$$

$$= \sum_{x' \in \mathcal{X}} \sum_{\substack{y \in \mathcal{Y} \\ q(y)=y'}} w(x')\delta_y(f(x + x'))$$

$$= \sum_{\substack{y \in \mathcal{Y} \\ q(y)=y'}} (\mathrm{LH}_w f)(x, y).$$

A.3   Proof of Theorem 3

The expected value of the local histogram (4.1) of a composite image (4.3) is:

$$\mathrm{E}_\Phi(\mathrm{LH}_w \mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x, y) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) \sum_{x' \in \mathcal{X}} w(x')\delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x + x')). \quad (\mathrm{A.3})$$

For any fixed φ, x, and x', we have $\varphi(x + x') = n$ for exactly one n. Thus, for any fixed x, x' and y, we can split a sum of $1_\varphi(x + x', n)\delta_y(f_n(x + x'))$ over all n into one summand where

$n = \varphi(x + x')$ and the remaining $N - 1$ summands for which $n \neq \varphi(x + x')$:

$$\sum_{n=0}^{N-1} 1_\varphi(x + x', n)\delta_y(f_n(x + x')) = (1)\delta_y(f_{\varphi(x+x')}(x + x')) + \sum_{n \neq \varphi(x+x')} (0)\delta_y(f_n(x + x'))$$

$$= \delta_y((\mathrm{occ}_\varphi\{f_n\}_{n=0}^{N-1})(x + x')), \tag{A.4}$$

where the final equality follows immediately from (4.3). Substituting (A.4) into (A.3) and using (4.5) yields:

$$\mathrm{E}_\Phi(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x, y) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) \sum_{x' \in \mathcal{X}} w(x') \left(\sum_{n=0}^{N-1} 1_\varphi(x + x', n)\delta_y(f_n(x + x')) \right)$$

$$= \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x')\delta_y(f_n(x + x')) \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) 1_\varphi(x + x', n)$$

$$= \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x')\delta_y(f_n(x + x'))\overline{1}_\Phi(x + x', n). \tag{A.5}$$

Rewriting (A.5) in terms of $\varepsilon := \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x')\delta_y(f_n(x + x'))[\overline{1}_\Phi(x + x', n) - \overline{1}_\Phi(x, n)]$ gives our first claim (4.6):

$$\mathrm{E}_\Phi(\mathrm{LH}_w\mathrm{occ}_\Phi\{f_n\}_{n=0}^{N-1})(x, y) = \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x')\delta_y(f_n(x + x'))\overline{1}_\Phi(x, n) + \varepsilon$$

$$= \sum_{n=0}^{N-1} \overline{1}_\Phi(x, n) \sum_{x' \in \mathcal{X}} w(x')\delta_y(f_n(x + x')) + \varepsilon$$

$$= \sum_{n=0}^{N-1} \overline{1}_\Phi(x, n)(\mathrm{LH}_w f_n)(x, y) + \varepsilon.$$

For the second claim, the fact that $|\delta_y(f_n(x + x'))| \leq 1$ and the triangle inequality bound on ε leads to:

$$|\varepsilon| = \left| \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') \delta_y(f_n(x + x')) [\overline{1}_\Phi(x + x', n) - \overline{1}_\Phi(x, n)] \right|$$

$$\leq \sum_{n=0}^{N-1} \sum_{x' \in \mathcal{X}} w(x') |\overline{1}_\Phi(x + x', n) - \overline{1}_\Phi(x, n)|.$$

Finally, to prove our third claim (4.7), note that for any fixed $x \in \mathcal{X}$, (4.5) gives:

$$\sum_{n=0}^{N-1} \overline{1}_\Phi(x, n) = \sum_{n=0}^{N-1} \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) 1_\varphi(x, n) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) \sum_{n=0}^{N-1} \begin{cases} 1, & \varphi(x) = n, \\ 0, & \varphi(x) \neq n. \end{cases} \quad (\text{A.6})$$

Since as previously noted we have $\varphi(x) = n$ for exactly one n, (A.6) becomes:

$$\sum_{n=0}^{N-1} \overline{1}_\Phi(x, n) = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) = 1.$$

A.4   Proof of Theorem 4

If Φ is flat, $\overline{1}_\Phi(x + x', n) = \overline{1}_\Phi(x, n)$ for all $x, x' \in \mathcal{X}$. The error bound in Theorem 3 then gives $\varepsilon = 0$. Denoting $\overline{1}_\Phi(x, n)$ as λ_n in (4.7) thus yields our claim.

A.5   Proof of Theorem 5

We begin by placing an equivalence relation $\sim$ on $\ell(\mathcal{X}, \mathbb{Z}_N)$, letting $\varphi' \sim \varphi$ when there exists some $x \in \mathcal{X}$ such that $\varphi' = \mathrm{T}^x \varphi$. Letting $\mathcal{R}$ denote a set of representatives from the corresponding equivalence classes, we have that for all $\varphi' \in \ell(\mathcal{X}, \mathbb{Z}_N)$, there exists a unique

$\varphi \in \mathcal{R}$ such that $\varphi' = \mathrm{T}^x \varphi$. As such,

$$\overline{1}_\Phi = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_N)} \mathrm{P}_\Phi(\varphi) 1_\varphi = \sum_{\varphi \in \mathcal{R}} \sum_{\varphi' \sim \varphi} \mathrm{P}_\Phi(\varphi') 1_{\varphi'}. \tag{A.7}$$

Now, fix any $\varphi \in \mathcal{R}$, and consider the subgroup $\mathcal{W}_\varphi = \{x \in \mathcal{X} : \mathrm{T}^x \varphi = \varphi\}$ of the finite abelian group $\mathcal{X}$. Letting $\mathcal{X}/\mathcal{W}_\varphi$ denote a fixed set of coset representatives of $\mathcal{X}$ with respect to $\mathcal{W}_\varphi$, we claim that $\beta : \mathcal{X}/\mathcal{W}_\varphi \to \{\varphi' : \varphi' \sim \varphi\}$, $\beta(x) := \mathrm{T}^x \varphi$ is a bijection.

Indeed, to show β is one-to-one, note that if $\mathrm{T}^x \varphi = \beta(x) = \beta(x') = \mathrm{T}^{x'} \varphi$, then $\mathrm{T}^{x-x'} \varphi = \varphi$, implying $x - x' \in \mathcal{W}_\varphi$; since x and x' are both coset representatives of $\mathcal{X}/\mathcal{W}_\varphi$, this is a contradiction unless $x = x'$. Meanwhile, to show β is onto, take any $\varphi' \sim \varphi$, and consider a corresponding x' such that $\varphi' = \mathrm{T}^{x'} \varphi$. Taking the unique $x \in \mathcal{X}/\mathcal{W}_\varphi$ and $w \in \mathcal{W}_\varphi$ such that $x' = x + w$, we have: $\varphi' = \mathrm{T}^{x'} \varphi = \mathrm{T}^{x+w} \varphi = \mathrm{T}^x (\mathrm{T}^w \varphi) = \mathrm{T}^x \varphi = \beta(x)$.

Invoking this claim, along with the assumed translation-invariance of Φ, yields:

$$\sum_{\varphi' \sim \varphi} \mathrm{P}_\Phi(\varphi') 1_{\varphi'} = \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} \mathrm{P}_\Phi(\beta(x)) 1_{\beta(x)}$$

$$= \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} \mathrm{P}_\Phi(\mathrm{T}^x \varphi) 1_{\mathrm{T}^x \varphi}$$

$$= \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} \mathrm{P}_\Phi(\varphi) 1_{\mathrm{T}^x \varphi}$$

$$= \mathrm{P}_\Phi(\varphi) \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} 1_{\mathrm{T}^x \varphi}. \tag{A.8}$$

Again, writing any $x' \in \mathcal{X}$ as $x' = x + w$, where $x \in \mathcal{X}/\mathcal{W}_\varphi$ and $w \in \mathcal{W}_\varphi$, gives:

$$\sum_{x' \in \mathcal{X}} 1_{\mathrm{T}^{x'} \varphi} = \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} \sum_{w \in \mathcal{W}_\varphi} 1_{\mathrm{T}^{x+w} \varphi} = \left(\sum_{w \in \mathcal{W}_\varphi} 1 \right) \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} 1_{\mathrm{T}^x \varphi} = |\mathcal{W}_\varphi| \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} 1_{\mathrm{T}^x \varphi}. \tag{A.9}$$

Substituting (A.9) into (A.8) gives:

$$\sum_{\varphi' \sim \varphi} \mathrm{P}_\Phi(\varphi') 1_{\varphi'} = \mathrm{P}_\Phi(\varphi) \sum_{x \in \mathcal{X}/\mathcal{W}_\varphi} 1_{\mathrm{T}^x \varphi} = \frac{\mathrm{P}_\Phi(\varphi)}{|\mathcal{W}_\varphi|} \sum_{x' \in \mathcal{X}} 1_{\mathrm{T}^{x'} \varphi}. \qquad (A.10)$$

Since $\displaystyle\sum_{x' \in \mathcal{X}} 1_{\mathrm{T}^{x'} \varphi}(x, n) = \sum_{x' \in \mathcal{X}} \begin{Bmatrix} 1, & \varphi(x - x') = n \\ 0, & \varphi(x - x') \neq n \end{Bmatrix} = \big|\{x' \in \mathcal{X} : \varphi(x') = n\}\big| = |\varphi^{-1}\{n\}|,$

substituting (A.10) into (A.7) gives:

$$\overline{1}_\Phi(x, n) = \sum_{\varphi \in \mathcal{R}} \sum_{\varphi' \sim \varphi} \mathrm{P}_\Phi(\varphi') 1_{\varphi'}(x, n) = \sum_{\varphi \in \mathcal{R}} \frac{\mathrm{P}_\Phi(\varphi)}{|\mathcal{W}_\varphi|} \sum_{x' \in \mathcal{X}} 1_{\mathrm{T}^{x'} \varphi}(x, n) = \sum_{\varphi \in \mathcal{R}} \frac{\mathrm{P}_\Phi(\varphi)}{|\mathcal{W}_\varphi|} |\varphi^{-1}\{n\}|,$$

implying Φ is flat, since the value of $\overline{1}_\Phi(x, n)$ depends only on n and is independent of x.

A.6 Proof of Theorem 6

We first show that (4.12) defines a probability density function, namely that values of $\mathrm{P}_{\Phi \star \Psi}(\sigma)$ over all σ in $\ell(\mathcal{X}, \mathbb{Z}_2)$ sum to 1. Since P_Φ is a probability density function by assumption, we have:

$$1 = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Phi(\varphi). \qquad (A.11)$$

Similarly, for any fixed $x \in \mathcal{X}$, we have:

$$1 = \sum_{\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Psi(\psi_x), \qquad (A.12)$$

where the subscript "x" on ψ indicates that this particular ψ is intended to expand φ at the particular point x as opposed to at some other point. Taking the product of (A.11) with the

product of (A.12) over all x yields:

$$1 = 1(1)^{|\mathcal{X}|} = \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} \sum_{\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_\Psi(\psi_x) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X}}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x),$$

$$(A.13)$$

where the final quantity in (A.13) contains all of the cross terms resulting from distributing the product over all sums of the form (A.12). Now, since for each choice of φ and $\{\psi_x\}_{x \in \mathcal{X}}$ there is exactly one resulting $\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$, we can rewrite (A.13) in terms of the definition (4.12) of $P_{\Phi \star \Psi}$, obtaining our claim:

$$1 = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sigma}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x) = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} P_{\Phi \star \Psi}(\sigma).$$

Thus, (4.12) indeed defines a probability density function, as claimed.

We next show that the occlusion model $\Phi \star \Psi$ is translation-invariant, if Φ is translation-invariant. To do this, we claim that if $T^{\tilde{x}}\sigma = \varphi \star \{\psi_x\}_{x \in \mathcal{X}}$ then $\sigma = (T^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}}$. To see this claim, note that

$$\sigma(x - \tilde{x}) = (T^{\tilde{x}}\sigma)(x) = (\varphi \star \{\psi_{x'}\}_{x' \in \mathcal{X}})(x) = 1$$

if and only if there exists some x', x'' in $\mathcal{X}$ such that $x = x' + x''$, $\varphi(x') = 1$, and $\psi_{x'}(x'') = 1$. Letting $\hat{x} = x - \tilde{x}$, we thus have that $\sigma(\hat{x}) = 1$ if and only if $\hat{x} = (x' - \tilde{x}) + x''$, where $(T^{-\tilde{x}}\varphi)(x' - \tilde{x}) = \varphi(x' - \tilde{x} + \tilde{x}) = \varphi(x') = 1$ and $\psi_{(x'-\tilde{x})+\tilde{x}}(x'') = \psi_{x'}(x'') = 1$, implying $\sigma = (T^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}}$, as claimed. Having the claim, (4.12) implies:

$$P_{\Phi \star \Psi}(T^{\tilde{x}}\sigma) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = T^{\tilde{x}}\sigma}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^\mathcal{X} \\ (T^{-\tilde{x}}\varphi) \star \{\psi_{x+\tilde{x}}\}_{x \in \mathcal{X}} = \sigma}} P_\Phi(\varphi) \prod_{x \in \mathcal{X}} P_\Psi(\psi_x).$$

To continue, we make the change of variables $\varphi' := \mathrm{T}^{-\tilde{x}}\varphi$ and $\psi'_x := \psi_{x+\tilde{x}}$:

$$\mathrm{P}_{\Phi\star\Psi}(\mathrm{T}^{\tilde{x}}\sigma) = \sum_{\substack{\varphi'\in\ell(\mathcal{X},\mathbb{Z}_2) \\ \{\psi'_x\}_{x\in\mathcal{X}}\in[\ell(\mathcal{X},\mathbb{Z}_2)]^{\mathcal{X}} \\ (\varphi')\star\{\psi'_x\}_{x\in\mathcal{X}}=\sigma}} \mathrm{P}_{\Phi}(\mathrm{T}^{\tilde{x}}\varphi') \prod_{x\in\mathcal{X}} \mathrm{P}_{\Psi}(\psi'_{x-\tilde{x}}).$$

Since Φ is translation-invariant and $\prod_{x\in\mathcal{X}} \mathrm{P}_{\Psi}(\psi'_{x-\tilde{x}}) = \prod_{x\in\mathcal{X}} \mathrm{P}_{\Psi}(\psi'_x)$, we have:

$$\mathrm{P}_{\Phi\star\Psi}(\mathrm{T}^{\tilde{x}}\sigma) = \sum_{\substack{\varphi'\in\ell(\mathcal{X},\mathbb{Z}_2) \\ \{\psi'_x\}_{x\in\mathcal{X}}\in[\ell(\mathcal{X},\mathbb{Z}_2)]^{\mathcal{X}} \\ (\varphi')\star\{\psi'_x\}_{x\in\mathcal{X}}=\sigma}} \mathrm{P}_{\Phi}(\varphi') \prod_{x\in\mathcal{X}} \mathrm{P}_{\Psi}(\psi'_x) = \mathrm{P}_{\Phi\star\Psi}(\sigma),$$

and so $\Phi \star \Psi$ is indeed translation-invariant (4.10), as claimed.

For our final claim, we assume that Φ and Ψ are effectively disjoint (4.13) and that either Φ or Ψ is flat. To do so, it is helpful to characterize the flatness of an arbitrary occlusion model Φ from $\mathcal{X}$ to $\mathbb{Z}_2$ in terms of the corresponding function $\overline{\Phi} := \sum_{\varphi\in\ell(\mathcal{X},\mathbb{Z}_2)} \mathrm{P}_{\Phi}(\varphi)\varphi$. Indeed, for any $\varphi : \mathcal{X} \to \mathbb{Z}_2$, (4.4) may be rewritten as $1_{\varphi}(x,1) = \varphi(x)$ and so:

$$\overline{1}_{\Phi}(x,1) = \sum_{\varphi\in\ell(\mathcal{X},\mathbb{Z}_2)} \mathrm{P}_{\Phi}(\varphi)1_{\varphi}(x,1) = \sum_{\varphi\in\ell(\mathcal{X},\mathbb{Z}_2)} \mathrm{P}_{\Phi}(\varphi)\varphi(x) = \overline{\Phi}(x). \qquad \text{(A.14)}$$

In light of (A.14), we claim that Φ is flat if and only if $\overline{\Phi}$ is constant. Indeed, if Φ is flat, then there exists λ_1 such that $\overline{\Phi}(x) = \overline{1}_{\Phi}(x,1) = \lambda_1$ for all $x \in \mathcal{X}$. Conversely, if $\overline{\Phi}(x)$ is constant, then there exists λ_1 such that $\overline{1}_{\Phi}(x,1) = \overline{\Phi}(x) = \lambda_1$ for all $x \in \mathcal{X}$; by (4.7), this further implies that $\overline{1}_{\Phi}(x,0) = 1 - \overline{1}_{\Phi}(x,1) = 1 - \lambda_1$ for all $x \in \mathcal{X}$ and so Φ is flat.

Having this claim, we show that $\Phi \star \Psi$ is flat by showing that $\overline{\Phi \star \Psi}$ is constant. To do this, we show that if Φ and Ψ are effectively disjoint then $\overline{\Phi \star \Psi} = \overline{\Phi} * \overline{\Psi}$ where "$*$" denotes

standard convolution over $\mathcal{X}$. According to the definition of $\Phi \star \Psi$ (4.12) we have:

$$\overline{\Phi \star \Psi} = \sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_{\Phi \star \Psi}(\sigma) \sigma$$

$$= \sum_{\substack{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)}} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}} \\ \varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sigma}} \mathrm{P}_\Phi(\varphi) \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) (\varphi \star \{\psi_x\}_{x \in \mathcal{X}}). \tag{A.15}$$

Since any particular choice of φ and $\{\psi_x\}_{x \in \mathcal{X}}$ produces a unique σ via $\star$ we can simplify (A.15) to

$$\overline{\Phi \star \Psi} = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}}} \mathrm{P}_\Phi(\varphi) \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) (\varphi \star \{\psi_x\}_{x \in \mathcal{X}}). \tag{A.16}$$

Moreover, since Φ and Ψ are effectively disjoint (4.13) we have $\varphi \star \{\psi_x\}_{x \in \mathcal{X}} = \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'} \psi_{x'}$

meaning (A.16) becomes:

$$\overline{\Phi \star \Psi} = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}}} \mathrm{P}_\Phi(\varphi) \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) \left(\sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'} \psi_{x'} \right)$$

$$= \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Phi(\varphi) \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'} \left[\sum_{\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}} \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) \psi_{x'} \right]. \tag{A.17}$$

Now, for any fixed $x' \in \mathcal{X}$ such that $\varphi(x') = 1$, we factor the corresponding innermost sum in (A.17) into a product of $|\mathcal{X}|$ distinct sums—one for each $x \in \mathcal{X}$—to obtain:

$$\sum_{\{\psi_x\}_{x \in \mathcal{X}} \in [\ell(\mathcal{X}, \mathbb{Z}_2)]^{\mathcal{X}}} \left(\prod_{x \in \mathcal{X}} \mathrm{P}_\Psi(\psi_x) \right) \psi_{x'} = \left[\prod_{x \neq x'} \left(\sum_{\psi_x \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Psi(\psi_x) \right) \right] \sum_{\psi_{x'} \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Psi(\psi_{x'}) \psi_{x'}$$

$$= \left(\prod_{x \neq x'} 1 \right) \overline{\Psi}$$

$$= \overline{\Psi}. \tag{A.18}$$

Substituting (A.18) into (A.17) then gives:

$$
\begin{aligned}
\overline{\Phi \star \Psi} &= \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Phi(\varphi) \sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \mathrm{T}^{x'} \overline{\Psi} \\
&= \sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Phi(\varphi) \left(\sum_{\substack{x' \in \mathcal{X} \\ \varphi(x')=1}} \delta_{x'} \right) * \overline{\Psi} \\
&= \left(\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Phi(\varphi)\varphi \right) * \overline{\Psi} \\
&= \overline{\Phi} * \overline{\Psi}.
\end{aligned}
$$

Thus, the effective disjointness of Φ and Ψ indeed implies $\overline{\Phi \star \Psi} = \overline{\Phi} * \overline{\Psi}$. As such, if we further assume that either Φ or Ψ is flat, then either $\overline{\Phi}$ or $\overline{\Psi}$ is constant, implying in either case that $\overline{\Phi \star \Psi}$ is constant and so $\Phi \star \Psi$ is flat.

A.7 Proof of Theorem 7

To show that (4.15) defines a probability density function on $\ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})$, note that:

$$
\begin{aligned}
1 &= (1)(1)(1) \\
&= \left(\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi})} \mathrm{P}_\Phi(\varphi) \right) \left(\sum_{\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi})} \mathrm{P}_\Psi(\psi) \right) \left(\sum_{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)} \mathrm{P}_\Sigma(\sigma) \right) \\
&= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2)}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma). \quad\quad (A.19)
\end{aligned}
$$

Noting that for each fixed φ, ψ, and σ, there exists exactly one $\upsilon \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})$ such that $\varphi \#_\sigma \psi = \upsilon$, (A.19) becomes:

$$1 = \sum_{\upsilon \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \varphi \#_\sigma \psi = \upsilon}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma) = \sum_{\upsilon \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi})} \mathrm{P}_{\Phi \#_\Sigma \Psi}(\upsilon),$$

as claimed. For the second conclusion, assume that Φ, Ψ, and Σ are flat. Our goal is to show that $\Phi \#_\Sigma \Psi$ is flat (4.8), meaning that for any $n \in \mathbb{Z}_{N_\varphi + N_\psi}$, we want to show that there exists a scalar λ_n such that:

$$\sum_{\substack{\upsilon \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ \upsilon(x) = n}} \mathrm{P}_{\Phi \#_\Sigma \Psi}(\upsilon) = \lambda_n \tag{A.20}$$

for all $x \in \mathcal{X}$. To see this, note that for any such x and n, we have:

$$\sum_{\substack{\upsilon \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ \upsilon(x) = n}} \mathrm{P}_{\Phi \#_\Sigma \Psi}(\upsilon) = \sum_{\substack{\upsilon \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ \upsilon(x) = n}} \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \varphi \#_\sigma \psi = \upsilon}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma)$$

$$= \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ (\varphi \#_\sigma \psi)(x) = n}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma). \tag{A.21}$$

Now, in the special case where $n = 0, \ldots, N_\varphi - 1$, (4.14) gives that $(\varphi \#_\sigma \psi)(x) = n$ if and only if $\varphi(x) = n$ and $\sigma(x) = 0$. As such, in this case (A.21) becomes:

$$\sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} \mathrm{P}_{\Phi \#_\Sigma \Psi}(v) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}), \varphi(x) = n \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2), \sigma(x) = 0}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma)$$

$$= \left(\sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \varphi(x) = n}} \mathrm{P}_\Phi(\varphi) \right) \left(\sum_{\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi})} \mathrm{P}_\Psi(\psi) \right) \left(\sum_{\substack{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \sigma(x) = 0}} \mathrm{P}_\Sigma(\sigma) \right)$$

$$= \lambda_{\Phi, n} \lambda_{\Sigma, 0}. \tag{A.22}$$

If, on the other hand $n = N_\varphi, \ldots, N_\varphi + N_\psi - 1$ then (4.14) gives that $(\varphi \#_\sigma \psi)(x) = n$ if and only if $\psi(x) = n - N_\varphi$ and $\sigma(x) = 1$. In this case, (A.21) becomes:

$$\sum_{\substack{v \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi + N_\psi}) \\ v(x) = n}} \mathrm{P}_{\Phi \#_\Sigma \Psi}(v) = \sum_{\substack{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi}) \\ \psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}), \psi(x) = n - N_\varphi \\ \sigma \in \ell(\mathcal{X}, \mathbb{Z}_2), \sigma(x) = 1}} \mathrm{P}_\Phi(\varphi) \mathrm{P}_\Psi(\psi) \mathrm{P}_\Sigma(\sigma)$$

$$= \left(\sum_{\varphi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\varphi})} \mathrm{P}_\Phi(\varphi) \right) \left(\sum_{\substack{\psi \in \ell(\mathcal{X}, \mathbb{Z}_{N_\psi}) \\ \psi(x) = n - N_\varphi}} \mathrm{P}_\Psi(\psi) \right) \left(\sum_{\substack{\sigma \in \ell(\mathcal{X}, \mathbb{Z}_2) \\ \sigma(x) = 1}} \mathrm{P}_\Sigma(\sigma) \right)$$

$$= \lambda_{\Psi, n - N_\varphi} \lambda_{\Sigma, 1}. \tag{A.23}$$

Thus, for any $x \in \mathcal{X}$ we either have (A.22) or (A.23) meaning $\Phi \#_\Sigma \Psi$ is flat (A.20), as claimed.

Bibliography

[1] L. C. Seeff, T. B. Richards, J. A. Shapiro, M. R. Nadel, D. L. Manninen, L. S. Given, F. B. Dong, L. D. Winges, and M. T. McKenna, "How many endoscopies are performed for colorectal cancer screening? Results from CDC's survey of endoscopic capacity," *Gastroenterology*, vol. 127, no. 6, pp. 1670 – 1677, 2004.

[2] The Intersociety Committee on Pathology Information, "Career opportunities in pathology," http://www.pathologytraining.org/recruit.ppt, 2011.

[3] National Center for Health Statistics Center for Disease Control and Prevention, "US mortality data 2007," 2010.

[4] American Cancer Society, "Cancer facts and figures 2010," http://www.cancer.org/Research/CancerFactsFigures/ CancerFactsFigures/cancer-facts-and-figures-2010.

[5] Statistical Research and Application Branch National Cancer Institute, "Devcan: Probability of developing or dying of cancer software, version 6.4.0." [Online]. Available: http://srab.cancer.gov/devcan

[6] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Paris, France, May 2008, pp. 496–499.

[7] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Paris, France, May 2008, pp. 284–287.

[8] S. Doyle, C. Rodriguez, A. Madabhushi, J. Tomaszeweski, and M. Feldman, "Detecting prostatic adenocarcinoma from digitized histology using a multi-scale hierarchical classification approach," in *Proc. IEEE Int. Conf. EMBS Soc.*, Aug. 2006, pp. 4759–4762.

[9] B. Karacali, A. Vamvakidou, and A. Tozeren, "Automated recognition of cell phenotypes in histology images based on membrane- and nuclei-targeting biomarkers," *BMC Medical Imaging*, vol. 7, no. 7, pp. 1–14, 2007. [Online]. Available: http://www.biomedcentral.com/1471-2342/7/7

[10] S. Petushi, F. Garcia, M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC Medical Imaging*, vol. 6, no. 1, p. 14, 2006. [Online]. Available: http://www.biomedcentral.com/1471-2342/6/14

[11] S. Keenan, J. Diamond, W. McCluggage, H. Bharucha, D. Thompson, P. Bartels, and P. Hamilton, "An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN)," *Journ. Pathology*, vol. 192, no. 3, pp. 351–362, 2000. [Online]. Available: http://www3.interscience.wiley.com/Journal/72514938/abstract

[12] M. Mete, X. Xu, C. Fan, and G. Shafirstein, "Head and neck cancer detection in histopathological slides," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2006, pp. 223–230.

[13] ——, "Automatic delineation of malignancy in histopathological head and neck slides," *BMC Bioinformatics*, vol. 8, no. Suppl 7, p. S17, 2007. [Online]. Available: http://www.biomedcentral.com/1471-2105/8/S7/S17

[14] D. Weaver, D. Krag, E. Manna, T. Ashikaga, S. Harlow, and K. Bauer, "Comparison of pathologist-detected and automated computer-assisted image analysis detected sentinel lymph node micrometastases in breast cancer," *Modern Pathology*, vol. 16, no. 11, pp. 1159–1163, 2003. [Online]. Available: http://www.nature.com/modpathol/Journal/v16/n11/abs/3880899a.html

[15] D. Glostos, P. Spyridonos, P. Petalas, D. Cavouras, P. Ravazoula, P. Dadioti, I. Lekka, and G. Nikiforidis, "Computer-based malignancy grading of astrocytomas employing a support vector machine classifier, the WHO grading system and the regular hematoxylin-eosin diagnostic staining procedure," *Anal. Quant. Cytol. Histol.*, vol. 26, no. 2, pp. 77–83, 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15131894

[16] G. Begelman, M. Pechuk, E. Rivlin, and E. Sabo, "System for computer-aided multiresolution microscopic pathology diagnostics," in *Proc. IEEE Int. Conf. Comp. Vis. Syst.*, Jan. 2006, p. 16.

[17] A. Sims, M. Bennett, and A. Murray, "Image analysis can be used to detect spatial changes in the histopathology of pancreatic tumours," *Phys. in Medicine and Biology*, vol. 48, no. 13, p. N183, 2003. [Online]. Available: http://stacks.iop.org/0031-9155/48/i=13/a=401

[18] O. Sertel, J. Kong, U. Catalyurek, G. Lozanski, J. Saltz, and M. Gurcan, "Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading," *Journ. Sig.*

Proc. Systems, vol. 55, no. 1-3, pp. 169–183, 2009. [Online]. Available: http://www.springerlink.com/content/k242h86nk5364n88/

[19] A. Tutac, D. Racoceanu, T. Putti, W. Xiong, W. Leow, and V. Cretu, "Knowledge-guided semantic indexing of breast cancer histopathology images," in *Proc. Int. Conf. Biomed. Eng. and Informatics*, vol. 2, May 2008, pp. 107–112.

[20] J. Kong, H. Shimada, K. Boyer, J. Saltz, and M. Gurcan, "Image analysis for automated assessment of grade of neuroblastic differentiation," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Apr. 2007, pp. 61–64.

[21] D. F. Gleason, G. T. Mellinger, and Veterans Administration Cooperative Urological Research Group, "Prediction of prognosis for prostatic adenocarcinoma by combined histologic grading and clinical staging," *The Journal of Urology*, vol. 111, no. 1, pp. 58–64, Jan 1974.

[22] D. N. Louis, H. Ohgaki, O. D. Wiestler, and W. K. Cavenee, *WHO Classification of Tumours of the Central Nervous System*, 4th ed. World Health Organization, 2007.

[23] A. Wetzel, "Computational aspects of pathology image classification and retrieval," *Journ. Supercomputing*, vol. 11, no. 3, pp. 279–293, 1997. [Online]. Available: http://www.springerlink.com/content/k62146u205u73356/

[24] L. Zheng, A. Wetzel, J. Gilbertson, and M. Becich, "Design and analysis of a content-based pathology image retrieval system," *IEEE Trans. Inform. Tech. in Biomed.*, vol. 7, no. 4, pp. 249–255, Dec. 2003.

[25] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. Woosley, X. Guan, C. Schmitt, and N. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Chicago, IL, Jun. 2009, pp. 1107–1110.

[26] R. S. Sandler, J. E. Everhart, M. Donowitz, E. Adams, K. Cronin, C. Goodman, E. Gemmen, S. Shah, A. Avdic, and R. Rubin, "The burden of selected digestive diseases in the united states," *Gastroenterology*, vol. 122, no. 5, pp. 1500 – 1511, 2002.

[27] J. E. Everhart, Ed., *The Burden of Digestive Diseases in the United States.* National Institute of Diabetes and Digestive and Kidney Diseases, U.S. Dept of Health and Human Services, 2008, NIH Publication 096433.

[28] G. J. Klinker, S. A. Shafer, and T. Kanade, "A physical approach to color image understanding," *Int. Journ. Comp. Vis.*, vol. 4, no. 1, pp. 7 – 38, 1990.

[29] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognition*, vol. 29, no. 8, pp. 1233 – 1244, 1996.

[30] T. Gevers and A. W. M. Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval," pp. 102 – 119, Jan, 2000.

[31] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, pp. 786–804, 1979.

[32] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Patt. Anal. and Mach. Intelligence*, vol. 18, no. 8, pp. 837 – 842, Aug. 1996.

[33] S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Trans. Image Proc.*, vol. 11, no. 10, pp. 1160 – 1167, Oct. 2002.

[34] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Patt. Anal. and Mach. Intelligence*, vol. 24, pp. 971–987, 2002.

[35] H. Blum and R. N. Nagel, "Shape description using weighted symmetric axis features," *Pattern Recognition*, vol. 10, no. 3, pp. 167 – 180, 1978.

[36] R. Malladi, J. A. Sethian, and B. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Patt. Anal. and Mach. Intelligence*, vol. 17, no. 2, pp. 158–175, Feb. 1995.

[37] C. Xu and J. L. Prince, "Snakes, shapes and gradient vector flow," *IEEE Trans. Med. Imag.*, vol. 7, pp. 359–369, Mar. 1998.

[38] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010.

[39] P. Viola and M. J. Jones, "Robust real-time face detection," vol. 57, no. 2, pp. 137 – 154, 2004.

[40] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comp. Vis.*, vol. 2, 1999, pp. 1150 – 1157.

[41] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

[42] I. T. Jolliffe, *Principal Component Analysis.* John Wiley & Sons, 2002.

[43] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," in *Proc. IEEE Int. Workshop Neur. Netw. for Signal Proc.*, Aug. 1999, pp. 41–48.

[44] R. Duda, P. Hart, and D. Stork, *Pattern Classification.* Englewood Cliffs, NJ: John Wiley & Sons, 2001.

[45] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[46] R. J. Schalkoff, *Artificial Neural Networks*, ser. Computer Science. McGraw-Hill, 1997.

[47] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. Springer, 2007.

[48] E. Parzen, "On estimation of a probability density function and mode," *Ann. of Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[49] A. Chebira, J. A. Ozolek, C. A. Castro, W. G. Jenkinson, M. Gore, R. Bhagavatula, I. Khaimovich, S. E. Ormon, C. S. Navara, M. Sukhwani, K. E. Orwig, A. Ben-Yehudah, G. Schatten, G. K. Rohde, and J. Kovačević, "Multiresolution identification of germ layer components in teratomas derived from human and nonhuman primate embryonic stem cells," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Paris, France, May 2008, pp. 979–982.

[50] A. Chebira, Y. Barbotin, C. Jackson, T. E. Merryman, G. Srinivasa, R. F. Murphy, and J. Kovačević, "A multiresolution approach to automated classification of protein subcellular location images," *BMC Bioinformatics*, vol. 8, no. 210, 2007.

[51] R. A. Kellogg, D. Delubac, A. Chebira, J. Kovačević, and S. F. Zappe, "Imaging technologies for high-throughput Drosophila functional genomics screens," in *Proc. BMES Annual Fall Meeting*, St. Louis, MO, Sep. 2008.

[52] A. Chebira, L. P. Coelho, A. Sandryhaila, S. Lin, W. G. Jenkinson, J. MacSleyne, C. Hoffman, P. Cuadra, C. Jackson, M. Püschel, and J. Kovačević, "An adaptive multiresolution approach to fingerprint recognition," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, San Antonio, TX, Sep. 2007, pp. 457 – 460.

[53] A. Chebira, "Adaptive multiresolution frame classification of biomedical images," Ph.D. dissertation, Carnegie Mellon Univ., 2008.

[54] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, ser. Signal Processing. Englewood Cliffs, NJ: Prentice Hall, 1995, http://waveletsandsubbandcoding.org/. [Online]. Available: http://waveletsandsubbandcoding.org/

[55] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Proc.*, vol. 2, no. 2, pp. 160–175, Apr. 1993.

[56] P. Hennings Yeomans, J. Thornton, J. Kovačević, and B. V. K. V. Kumar, "Wavelet packet correlation methods in biometrics," *Appl. Opt., sp. iss. Biometric Recogn. Systems*, vol. 44, no. 5, pp. 637–646, Feb. 2005.

[57] A. Hass, G. Matheron, and J. Serra, "Morphologie mathématique et granulométries en place," *Annales des Mines - Part I*, vol. 9, pp. 736–753, 1967.

[58] G. Matheron, *Random Sets and Integral Geometry*. New York, NY: John Wiley & Sons, 1975.

[59] J. Serra, *Image Analysis and Mathematical Morphology*. New York, NY: Academic Press, 1982.

[60] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," vol. 9, no. 4, pp. 532 – 550, July 1987.

[61] S. Beucher, "The watershed algorithm applied to image trasnformation," *Scanning Microscopy*, vol. 6, pp. 299–314, 1992.

[62] G. Matheron, *Elments pour une Thorie des Milieux Poreux*. Paris, France: Masson, 1967.

[63] D. Stoyan and K. Mecke, *The Boolean Model: from Matheron till Today*. New York, NY: Springer, 2005, pp. 151 – 181.

[64] G. Matheron, "Model sequential de partition aleatoire," CMM, Tech. Rep., 1968.

[65] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *Int. Journ. Comp. Vis.*, vol. 41, pp. 35–59, 2001.

[66] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed.   McGraw-Hill, 2001.

[67] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imaging Vision*, vol. 18, no. 1, pp. 17–33, Jan. 2003.

[68] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention," *Int. Journ. Comp. Vis.*, vol. 11, pp. 283 – 318, 1993.

[69] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123 – 140, 1996.

[70] ——, "Random forests," *Machine Learning*, vol. 45, pp. 5 – 32, 2001.

[71] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*.   Belmont, CA: Wadsworth, 1984.

[72] T. Lindeberg, *Scale-Space Theory in Computer Vision*.   Norwell, MA: Kluwer Academic Publishers, 1994.

[73] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proc. Int. Joint Conf. on Neural Networks*, vol. 1, Jun. 1989, pp. 593 – 605.

[74] L. Ficsor, V. Varga, A. Tagscherer, Z. Tulassay, and B. Molnar, "Automated classification of inflammation in colon histological sections based on digital microscopy and advanced image analysis," *Cytometry*, vol. 73A, pp. 230–237, 2008.

[75] C. Gunduz-Demir, M. Kandemir, A. Tosun, and C. Sokmensuer, "Automatic segmentation of colon glands using object-graphs," *Med. Image Anal.*, vol. 14, pp. 1–12, 2010.

[76] P. Huang, C.-H. Lee, and P.-L. Lin, "Support vector classification for pathological prostate images based on texture features of multi-categories," in *Proc. IEEE Int. Conf. Syst., Man and Cybern.*, 2009, pp. 912–916.

[77] H. Qureshi, O. Sertel, N. Rajpoot, R. Wilson, and M. Gurcan, "Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification," in *Proc. Int. Conf. Med. Image Computing and Computer-Assisted Intervention*, 2008, pp. 196–204.

[78] D. Gabor, "Theory of communication," *Journ. IEE*, vol. 93, pp. 429–457, 1946.

[79] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *Journ. Neuropsychology*, vol. 58, pp. 1233–1258, 1987.

[80] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, pp. 1169–1179, 1988.

[81] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Patt. Recogn.*, vol. 24, pp. 1167–1186, 1991.

[82] J. Daugman, "How iris recognition works," *IEEE Trans. Circ. and Syst.*, vol. 14, pp. 21–30, 2004.

[83] J. Soares, J. Leandro, R. Cesar, H. Jelinek, and M. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, pp. 1214–1222, 2006.

[84] S. Mohamed and M. Salama, "Prostate cancer spectral multifeature analysis using trus images," *IEEE Trans. Med. Imag.*, vol. 27, pp. 548–556, 2008.

[85] D. Zhao, Y. Chen, and H. Correa, "Statistical categorization of human histological images," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 3, Sep. 2005, pp. 628–631.

[86] ——, "Automated classification of human histological images, a multiple-instance learning approach," in *Proc. IEEE Life Sci. Syst. and Appl. Workshop*, Jul. 2006, pp. 1–2.

[87] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proc. IEEE Int. Conf. Patt. Recogn.*, vol. 1, 1994, pp. 582–585.

[88] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Patt. Anal. and Mach. Intelligence*, vol. 28, pp. 2037–2041, 2006.

[89] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Patt. Anal. and Mach. Intelligence*, vol. 29, pp. 915–928, 2007.

[90] M. Hafner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vecsei, and F. Wrba, "Pit pattern classification using extended local binary patterns," in *Proc. Int. Conf. Inform. Tech. and Appl. in Biomedicine*, Nov. 2009, pp. 1–4.

[91] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Patt. Recogn.*, vol. 42, pp. 425–436, 2009.

[92] N. N. Kachouie and P. Fieguth, "A medical texture local binary pattern for trus prostate segmentation," in *Proc. IEEE Int. Conf. EMBS Soc.*, Aug. 2007, pp. 5605–5608.

[93] B. Li and M.-H. Meng, "Small bowel tumor detection for wireless capsule endoscopy images using textural features and support vector machine," in *Int. Conf. Intelligent Robots and Syst.*, Oct. 2009, pp. 498–503.

[94] H. R. Tavakoli, H. R. Pourreza, and S. R. Quchani, "Study of Gabor and local binary patterns for retinal image analysis," in *Int. Workghop Adv. Comp. Intelligence*, Aug. 2010, pp. 527–532.

[95] B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature*, vol. 290, pp. 91–97, Mar. 1981.

[96] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. Journ. Comp. Vis.*, vol. 43, pp. 29–44, 2001.

[97] J. Kaufhold, R. Collins, A. Hoogs, and P. Rondot, "Recognition and segmentation of scene content using region-based classification," in *Proc. IEEE Int. Conf. Patt. Recogn.*, vol. 1, 2006, pp. 755–760.

[98] C. Zhong, Z. Sun, and T. Tan, "Robust 3D face recognition using learned visual codebook," in *Proc. IEEE Int. Conf. Comp. Vis. and Patt. Recogn.*, Jun. 2007, pp. 1–6.

[99] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," *Proc. IEEE Int. Conf. Comp. Vis. and Patt. Recogn.*, pp. 1–8, 2008.

[100] A. Bosch, X. Munoz, A. Oliver, and J. Marti, "Modeling and classifying breast tissue density in mammograms," in *Proc. IEEE Int. Conf. Comp. Vis. and Patt. Recogn.*, vol. 2, 2006, pp. 1552–1558.

[101] P. Khurd, C. Bahlmann, P. Maday, A. Kamen, S. Gibbs-Strauss, E. Genega, and J. Frangioni, "Computer-aided Gleason grading of prostate cancer histopathological images using texton forests," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Apr. 2010, pp. 636–639.

[102] S. Chatzistergos, J. Stoitsis, A. Papaevangelou, G. Zografos, and K. Nikita, "Parenchymal breast density estimation with the use of statistical characteristics and textons," in *Proc. Int. Conf. Inform. Tech. and Appl. in Biomedicine*, Nov. 2010, pp. 1–4.

[103] Y. Freund and R. Schapire, "A short introduction to boosting," *Journ. Japanese Soc. Artif. Intelligence*, vol. 14, no. 5, pp. 771–780, 1999.

[104] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class svm for learning in image retrieval," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, 2001, pp. 34 – 37.

[105] C. Yuan and D. Casasent, "A novel support vector classifier with better rejection performance," in *Proc. IEEE Int. Conf. Comp. Vis. and Patt. Recogn.*, vol. 1, Jun. 2003, pp. 419 – 424.

[106] D. M. J. Tax and R. P. W. Duin, "Data domain description using support vectors," in *Proc. of the European Symp. on Artificial Neural Networks*, 1999, pp. 251 – 256.