CONTRIBUTIONS TO COMPUTER-AIDED DIAGNOSIS OF PROSTATE CANCER
IN HISTOPATHOLOGY

By

Kien Nguyen

A DISSERTATION

Submitted
to Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science – Doctor of Philosophy

2013

ABSTRACT

CONTRIBUTIONS TO COMPUTER-AIDED DIAGNOSIS OF PROSTATE CANCER IN HISTOPATHOLOGY

By

Kien Nguyen

Prostate cancer is one of the most common and dreadful type of cancer in men. Due to the unclear symptoms of the disease, the diagnosis of prostate cancer is difficult, and requires multiple procedures. Among these procedures, the most important one is the examination of the prostate tissue biopsy to detect the presence of cancer regions in the tissue, and assign a Gleason score to the tissue (which determines the severity of the cancer). The detection and grading processes are based on the glandular structures as well as the cytological properties of the tissue. In a traditional examination, pathologists have to look at the tissue biopsy under a microscope. With the developments in digital pathology, especially in virtual microscopy, glass tissue slides can be digitized to generate tissue images. These images can be displayed on a monitor, annotated by software tools, and forwarded to experts for examination and diagnosis. However, the large volume of tissue images that are generated poses a challenge for pathologists to efficiently and accurately perform the diagnosis. Hence, there is a need to develop tools for automatic processing of prostate tissue images, which can assist pathologists in decision making and improve the throughput. This thesis deals with design and development of automatic tools for processing and analyzing prostate tissue images.

In tissue examination, the grading of tissue slides is a standard procedure to determine the severity of cancer. The most popular grading method is the Gleason grading, which relies on the gland structures in the tissue to assign a Gleason score ranging from 2 to 10 to the tissue image. We utilize the Gleason grading method in automated systems by segmenting glands from the tissue image and extracting features to discriminate them. By thorough analyses and evaluations, we demonstrate that the proposed methods lead to better gland

classification accuracies than published methods in the literature.

We utilize the proposed gland segmentation and gland feature extraction methods to solve the tissue image classification problem, which receives the most attraction in the literature. By comparing with popular texture-based methods, we show that using the proposed gland features is a better solution for this problem. To further improve the Gleason grade 3 vs Gleason grade 4 classification result, we propose a different approach for gland segmentation and study the properties of the nuclei arrangement in the segmented glands.

When a medical laboratory technician or a medical student who is not very experienced with Gleason grading wants to gain additional experience in the grading process, it will be useful if there was an image retrieval engine that could search for tissue regions similar to the region of interest (ROI) in the tissue slides that were annotated by experienced pathologists. The technician (or student) can use the retrieved regions (whose Gleason grades are known) as the references to grade the ROI. To create such an image retrieval search engine, we develop a gland-based method to compute the similarity between two tissue regions.

Besides gland structure information, cytological features of the prostate tissue (which are not used in Gleason grading) also provide useful information for pathologists to detect cancer. Cytological features refer to size, shape, quantity, and arrangement of the basic elements of the tissue such as cell, cytoplasm, nuclei, and nucleoli. One of the most important cytological features is that in cancer glands, nuclei usually contain nucleoli, while nuclei in normal glands do not. To utilize this information, we present a novel method to identify the nuclei with prominent nucleoli (NwN) in the tissue. By applying the result of NwN identification in a prostate cancer detection problem, we see that the use of cytological feature, i.e., the presence of NwN in this case, helps to boost the cancer detection accuracy.

With the proposed solutions for different problems in automated prostate cancer grading presented in this thesis, we believe that we are able to provide pathologists with useful tools to assist them with the prostate cancer diagnosis task.

To my loving parents

ACKNOWLEDGEMENTS

During the five years in my PhD program, I receive valuable help and supports from great people around me. I would like to take this opportunity to express my sincere thanks to everyone.

First and foremost, I would like to express my deep gratitude to my advisor, Professor Anil K. Jain. Countless of great things have been said about Professor Jain. I cannot cover all of them here, but I can tell a small part of them. I still remember the first time Professor Jain called to the lab and asked me to join the project in digital pathology. It is a big luck for me to work in this project with him. The longer time I worked, the more I felt impressed with his generosity and professionalism towards students. Professor Jain always gives students the best chance to learn and work. To bring me to the new field of research, he sent me to a workshop, from which I started to learn about the exciting field of study, medical image processing. To help me improve the knowledge in computer vision, he let me take the summer school in Italy. To create the good collaboration with researchers from Ventana Medical Systems, Inc., he helped me get the internships at the company and sent me there several times during the breaks. He also invited Dr. Anindya Sarkar from Ventana to attend my comprehensive exam and the final PhD defense. When I got a paper accepted at a MICCAI workshop, he encouraged me to attend the main conference as well. Moreover, he gave me a research assistantship appointment every semester so that I can concentrate on the research. In all these events, he generously paid for them without a hesitation. In the last year of my PhD program, he strongly supported me to find a job. Especially, Professor Jain and Mrs. Jain always provide people in the lab with great food and delicious sweet.

There are hundreds of things that I need to learn from Professor Jain. One of the most important things is that when doing research, we need to care about writing, presentation and packaging as much as caring about developing the algorithms. A great idea will be

Sertel, who gave me useful advices and supports during the time I was at Ventana.

I would like to thank Professor George Stockman, who gave me a lot of advices and warm supports. I had wonderful time with him for the canoe trips, his parties and his sense of humour. He is the one who shows me the beauty of Michigan. I would like to thank Mrs. Linda Moore and Mrs. Norma Teague, who always help me to handle the requirements of the degree program and the travelling arrangements. I would like to thank Dr. Eric Torng, the graduate director, who helped me to obtain the great teaching experience. I would like to thank the PhD committee members Dr. Yiying Tong, Dr. Xiaoming Liu, Dr. Shantanu Chakrabartty for the valuable comments and questions regarding the thesis.

I would like to thank Serhat Selcuk Bucak and Radha Chitta, who are considered my mentors in the machine learning problems. The discussions with them are very useful for my research. I would like to thank other PRIP fellows: Dr. Abhishek Nagar, Dr. Brendan Klare, Dr. Hu Han, Dr. Eryun Liu, Dr. Kai Cao, Alessandra Paulino, Sunpreet Arora, Scott Klum, Lacey Best-Rowden, Charles Otto and Soweon Yoon for the research discussions as well as the entertaining conversations and the fun time we had together.

Last but not least, I would like to thank my Vietnamese friends (Hiep Chau, Dinh Pham, Do Dat, Lan Dang, Viet Ho, Tuan Nguyen, Thang Nguyen, Chinh Dang, Son Tran, Loan Cao, Trieu Le, Trung Nguyen, Truong Do, and many more) and the Vietnamese association with whom I share the fun and joy during the five years. They make me feel like living at home.

Thank you every one !

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

Chapter 1

INTRODUCTION

Cancer[1] has become a severe threat to human lives due to its prevalence. According to the American Cancer Society [1], the estimated number of new cases, for all types of cancer in the U.S. for the year 2013 is 1,660,290, and the estimated number of deaths from cancer is 580,350. In men, prostate cancer is the most prevalent cancer type (other prevalent cancers are lung, colorectum, and bronchus). The estimated new cases of prostate cancer is 238,590 (Table 1.1), which is the highest number among all types of cancer in men (accounting for 28% of all the new cancer cases). Furthermore, the estimated number of deaths from this type of cancer is 29,720 (Table 1.2), which is the second highest (after lung and bronchus).

The distribution of prostate cancer incidences around the world is illustrated in Figure 1.1. According to this map, we can see that prostate cancer is concentrated in North America, Australia, and Northern Europe. Detailed information about the prostate and prostate cancer can be found in [15] and [16]. In this chapter, we summarize information about prostate cancer extracted from these studies.

Prostate is a gland in the male body. As can be seen in Figure 1.2, the prostate is located next to the bladder. The function of the prostate is to create the fluid to protect the sperm cells. A commonly occurring problem in the prostate is "Benign Prostatic Hyperplasia", in which the size of the prostate increases as men get older. The enlarged prostate obstructs the urethra, which makes it difficult for passing urine. Despite having similar symptoms with prostate cancer, "Benign Prostatic Hyperplasia" is not considered a cancer [15]. Prostate cancer occurs mostly in the gland cells (although there are many other cell types present in

---

[1]Cancer, tumor, and carcinoma refer to the tissues which are detected to have malignant properties of a cancer (cells grow aggressively, invade the surrounding tissues and spread to the non-adjacent tissues). Normal, benign and nontumor refer to the tissues which are not cancerous.
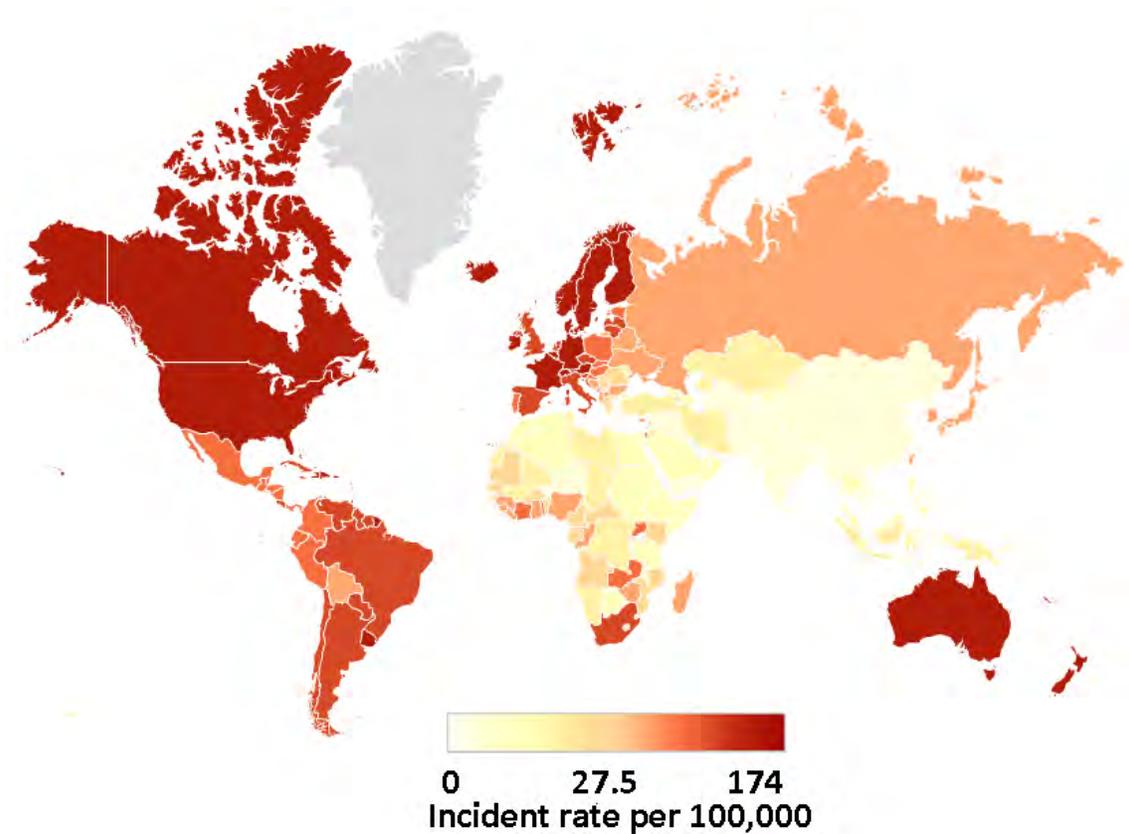
Figure 1.1: World map of prostate cancer incidence for the year 2008 [5]. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Cancer type	Number of estimated new cases	Percent
Prostate	**238,590**	**28%**
Lung & bronchus	118,080	14%
Colorectum	73,680	9%
Urinary bladder	54,610	6%
Melanoma of the skin	45,060	5%
Kidney & renal pelvis	40,430	5%
Non-Hodgkin lymphoma	37,600	4%
Oral cavity & pharynx	29,620	3%
Leukemia	27,880	3%
Pancreas	22,740	3%
All types	854,790	100%

Table 1.1: Estimated new cases of cancer in men in the U.S. for the year 2013 [1].

Cancer type	Number of deaths	Percent
Lung & bronchus	87,260	29%
Prostate	**29,720**	**9%**
Colon & rectum	26,300	9%
Pancreas	19,480	6%
Liver & intrahepatic bile duct	14,890	5%
Leukemia	13,660	4%
Esophagus	12,220	4%
Urinary bladder	10,820	3%
Non-Hodgkin lymphoma	10,590	3%
Kidney & renal pelvis	8,780	3%
All types	306,920	100%

Table 1.2: Estimated deaths due to cancer in men in the U.S. for the year 2013 [1].
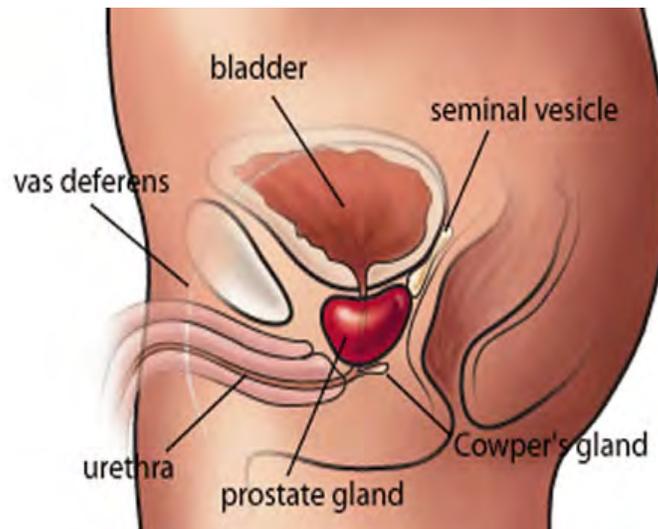
Figure 1.2: Prostate in the male body [6].

the prostate, cancer rarely occurs in those cells), thus, we call it *prostate adenocarcinoma* (*adenocarcinoma* refers to the cancer developing in gland cells). In this thesis, we only focus on prostate adenocarcinoma.

In most cases, prostate cancer grows rather slowly (e.g. many older men who have prostate cancer die due to other diseases even before prostate cancer afflicts them). Patients suffering from prostate cancer may not have any obvious symptoms or may have very few symptoms such as frequent urination, blood in the urine or painful urination. Prostate cancer can invade (metastasize) to different organs of the body such as rectum, lymph nodes, bones

3
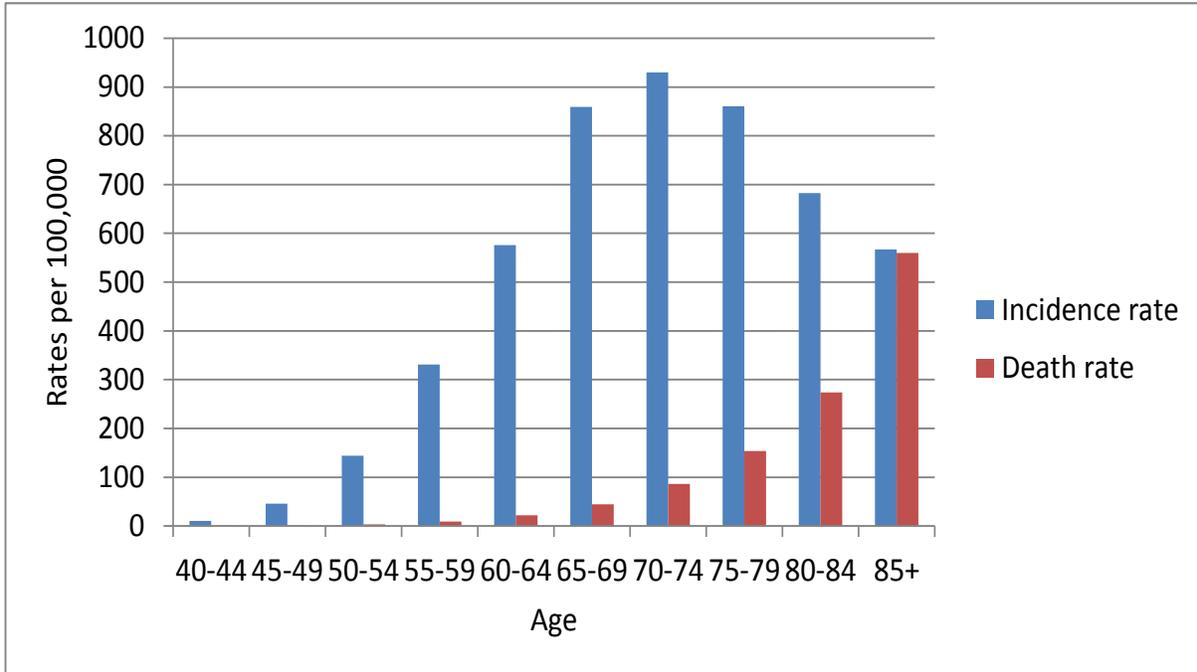
Figure 1.3: Prostate cancer incidence and death rate for men, as a function of age, in the U.S. during 2005-2009 [7].

and bladder. If cancer spreads to other organs, additional symptoms appear such as bone pain [15]. Although the cause of prostate cancer is not completely known, prostate cancer is believed to be related to family history and age. It is uncommon for prostate cancer to occur in men before the age of 45 but is more frequent for men over 50 (Figure 1.3). The risk for prostate cancer increases by two times for men whose father or brother had prostate cancer compared to men without a family history [17]. High blood pressure also increases the risk for prostate cancer [18]. Moreover, there is some tentative evidence for a correlation between dietary habits and prostate cancer [19]. Another study [20] found that consumption of processed meat and red meat may increase the incidence of prostate cancer. Some of the advice given to prevent prostate cancer includes: regular physical activity, eating vegetables (cabbage, cauliflower, broccoli, tomato, soy, beans) and fish, Vitamin E and mineral supplements [15].

The treatment methods prescribed for prostate cancer include: surgery, radiation therapy, active surveillance, cryosurgery, vaccine treatment, hormone therapy, and chemotherapy [15].

4

These treatments are usually undertaken individually. However, in some situations, these different treatments can be combined together. The diagnosis of prostate cancer involves rectal examination, blood test, biopsy tissue examination, and imaging test (more details about the diagnostic process will be discussed in the next section). Due to the availability of different imaging technologies and devices such as computed tomography (CT), magnetic resonance imaging (MRI), virtual microscopy (tissue scanner) (Figure 1.4), the amount of image data being generated increase significantly. This has lead to the development of computer-aided diagnosis (CAD) systems, which utilize image processing, pattern recognition, and machine learning algorithms to assist physicians in cancer diagnosis. In this thesis, we concentrate on the CAD systems which analyze histopathological images of the prostate, i.e. the digitized prostate tissue slides obtained from the biopsy. Specifically, we aim at developing automatic tools to assist pathologists in detecting cancer regions in the prostate tissue as well as in grading the tissue. These types of images are generated by a tissue slide scanner and capture cellular information in the prostate tissue. Before scanning, the tissue must be stained by some pigments to reveal the individual tissue components. Figure 1.5 shows an example of the histopathological image of the prostate tissue, which is stained by the well-known H&E method (explained in the subsequent sections).

The remainder of this chapter is organized as follows. In section 1.1, we summarize the prostate cancer diagnosis process performed by physicians, urologists, and pathologists. In section 1.2, we provide details of the cancer grading procedure performed by pathologists using images of prostate histopathology. Section 1.3 addresses the challenges in the automatic cancer grading systems. Section 1.4 briefly reports the prior work reported in the literature. Section 1.5 summarizes the contributions of the thesis. Section 1.6 describes the databases used in our experiments and section 1.7 describes the organization of the thesis.

(a) A computed tomography (CT) machine [21]

(b) A magnetic resonance imaging (MRI) machine [22]

(c) A tissue slide scanner [23]

Figure 1.4: Different imaging devices used in prostate cancer diagnosis.
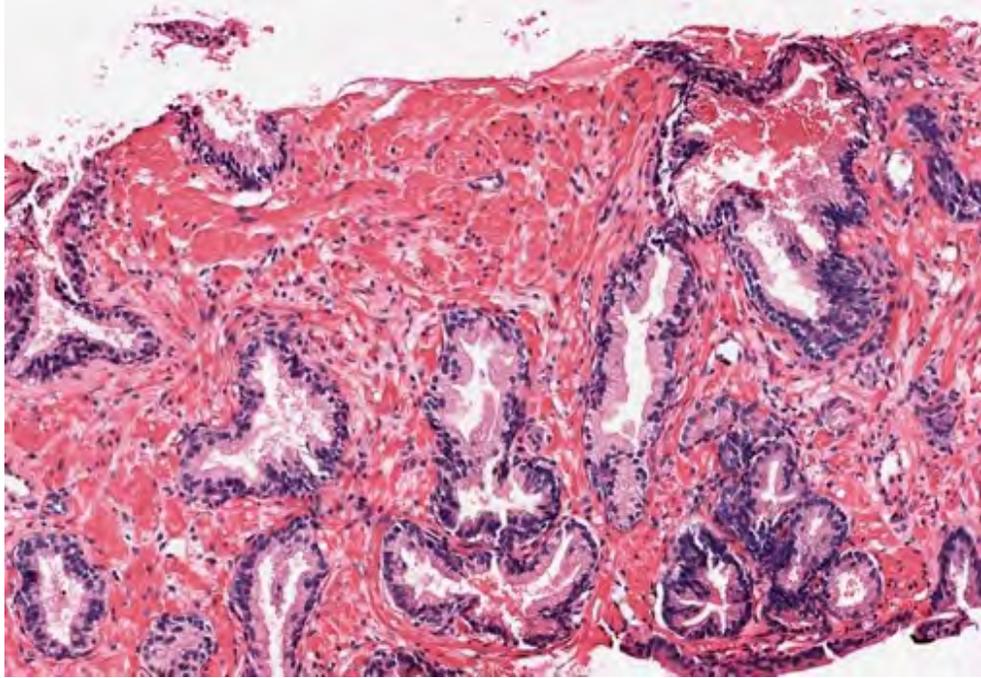
Figure 1.5: A prostate histopathological image. The tissue is stained by the H&E staining method. The pink areas denote stroma regions. The blue and purple areas denote the epithelial cell regions, including epithelial nuclei (blue) and epithelial cytoplasm (purple). The image size is 330 × 500 pixels, corresponding to a 0.66 × 1 mm^2 tissue region digitized at 5× magnification.

1.1 Prostate Cancer Diagnosis

The diagnostic process of prostate cancer [15] is delineated in Figure 1.6. The physician first asks the patient about his medical history and symptoms. Next, a prostate cancer screening is conducted. There are two options for the screening: a physical exam called digital rectal exam (DRE) and a prostate-specific antigen (PSA) blood test. In the DRE exam, the doctor directly inserts his finger into the patient's rectum to find hard areas in the prostate, which may be cancer areas. Besides the DRE, a PSA blood test [24] can also be conducted, mainly when the patient has no symptoms. This test aims to measure the PSA level in the blood of the patient (PSA is a substance produced by prostate gland cells, by both cancer and normal cells). In healthy men, the PSA level in blood is mostly below 4 ng/mL (nanograms per milliliter). The high PSA level indicates the possibility of prostate cancer. If the results

Medical
history and
symptoms

Digital
rectal
exam

Rectum

Prostate

PSA blood
test

Prostate
needle
biopsy

Histopathological
grading

Ultrasound,
CT scan,
MRI scan
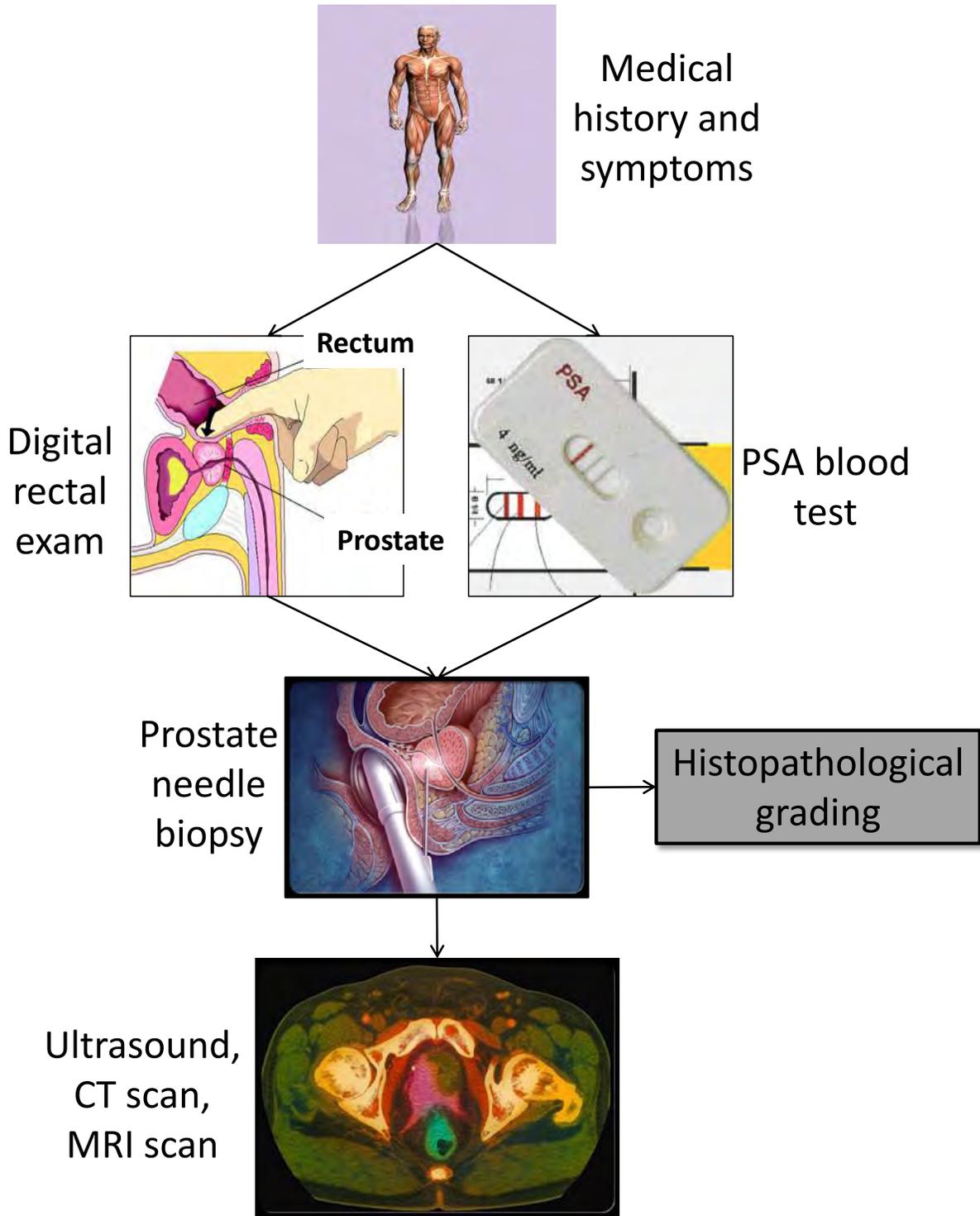
Figure 1.6: Prostate cancer diagnosis process [8–10].

of the DRE and/or PSA test suggest that the patient may have prostate cancer, a *prostate biopsy* is conducted. In this biopsy procedure, a urologist uses a needle to remove prostate tissue samples from the patient. Usually 12 tissue samples are removed. The tissue samples are examined under a microscope by a pathologist[2] to find cancer regions, and assign a grade to the cancer (which determines the malignancy level of the cancer). We refer to this process as *histopathological grading* process, which will be discussed in greater details in the next section. By combining the results of the DRE, PSA test and histopathological grading, the spread of the cancer is determined. If the cancer has spread beyond the prostate, additional imaging tests such as transrectal ultrasound (TRUS), bone scan, computed tomography (CT), magnetic resonance imaging (MRI) are conducted to find cancer in other organs of the body.

1.2   Tissue Preparation and Grading

The main focus of the thesis is the histopathological grading step in the prostate cancer diagnosis, i.e. the process of examining the tissue biopsy to detect cancer regions and grade the tissue (Figure 1.10). Before being graded, the tissue biopsy must undergo a preparation procedure.

1.2.1   Tissue Preparation and Digitization

After being removed from the patient's prostate, the tissue biopsy is processed with fixative to prevent it from becoming decayed. The tissue is then sectioned into slices by using a microtome (a machine which can create very thin slices), and are arranged on the glass slides before being stained. The staining procedure uses certain pigments to reveal the tissue components. Among the staining methods, H&E (Hematoxylin and Eosin combination) stain is the most widely used (the images we use in this thesis are images taken from the

---

[2]Pathologists are doctors who work with tissue samples to diagnose the disease.
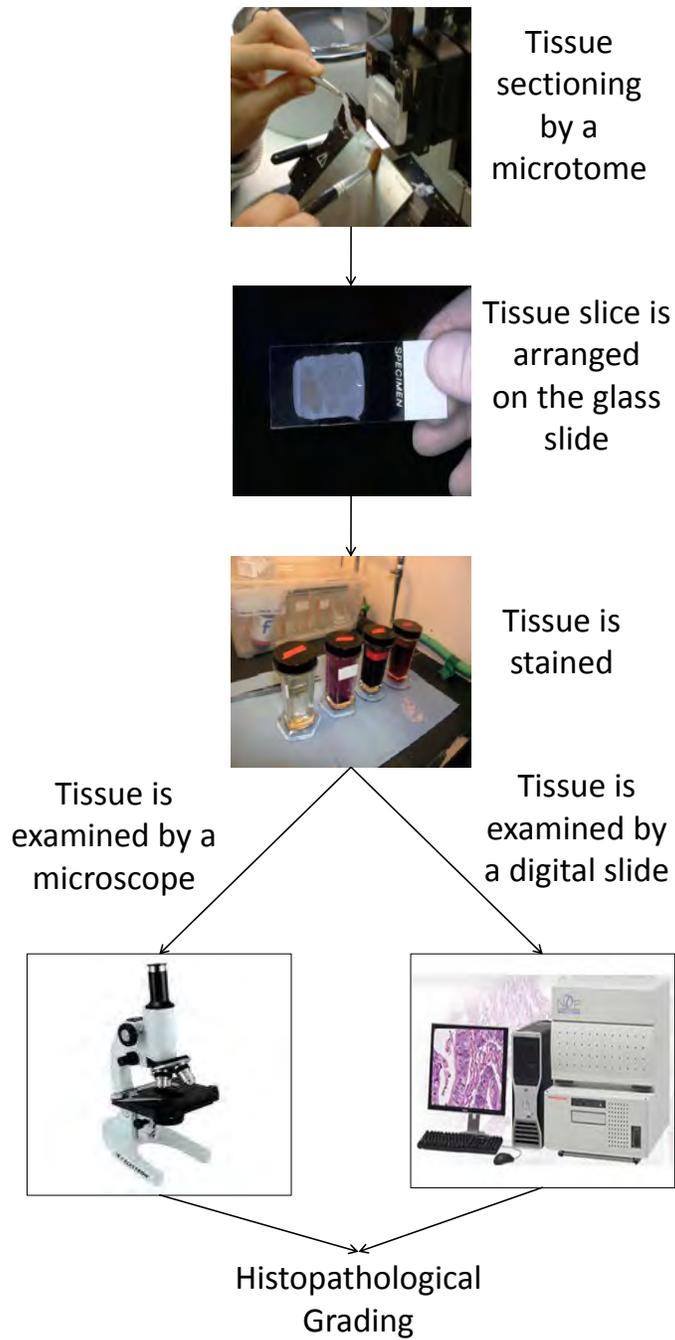
Figure 1.7: Procedure to create a tissue slide [11–13]. The slide is either examined directly under a microscope or digitized and processed on computer.

H&E stained tissues). Nuclei are stained blue by Hematoxylin, while cytoplasm and the tissue connective structures are stained pink. The pathologist then interprets the stained tissue. In the conventional method, tissue slides are examined using a microscope. However, recent developments in digital pathology, especially the development of virtual microscopy technique, allow us to create a digital image of the tissue slide (Figure 1.8) by using a digital tissue scanner (Figure 1.4c) shows the iScan Coreo Au slide scanner, manufactured by Ventana Medical Systems, Inc. [23]). Most of the available tissue slide scanners can digitize the tissue slides at $20\times$ or $40\times$ magnification [23, 25–27]. Only a few scanners can digitize tissue slides at a higher magnification, e.g. $100\times$ magnification [28, 29]. With the digital tissue image, pathologists can examine and manipulate the tissue image on a monitor. These digital slides can also be shared between pathologists to facilitate collaboration in grading difficult slides. Further, the online digital slides allow pathologists to work virtually anywhere. Figure 1.7 summarizes the tissue preparation and digitization process.

The volume of histopathological image data related to prostate tissue being produced is huge, which raises a challenge for pathologists. For example, an image representing a tissue area of 20×25 mm^2 at $20\times$ magnification contains $40{,}000 \times 50{,}000$ pixels, resulting in 6 GB of data. Moreover, multiple tissue sections are obtained for a single patient [30]. As a result, it is both time-consuming and tedious for pathologists to investigate the tissue structures and detect cancer regions from these large images. Moreover, there are problems with inter and intra-reader variability in reading the tissue slides [31–34], e.g. different pathologists may end up with different grading results for the tissue slide. Thus, there is a need to design and develop computer-aided systems to assist pathologists in detecting and grading the cancer in prostate tissue images.

1.2.2 Histopathological Grading

The goal of histopathological grading is to determine the severity of the cancer development in the tissue. Figure 1.10 summarizes the grading process performed by a pathologist. The
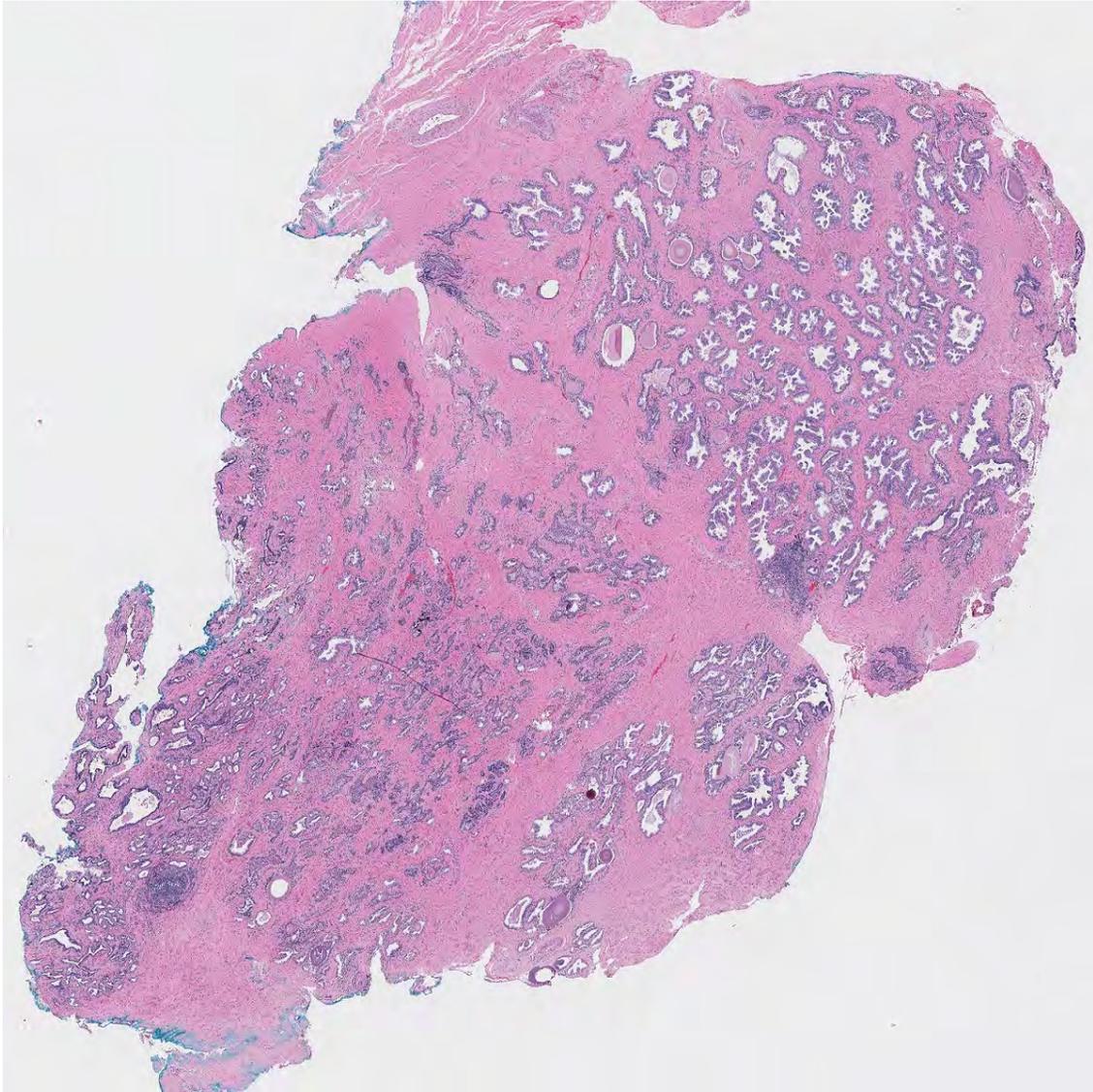
Figure 1.8: A digital image generated by a tissue slide scanner. The image size is 5,000 × 4,500 pixels, corresponding to a 10×9 mm^2 tissue slide digitized at 5× magnification.
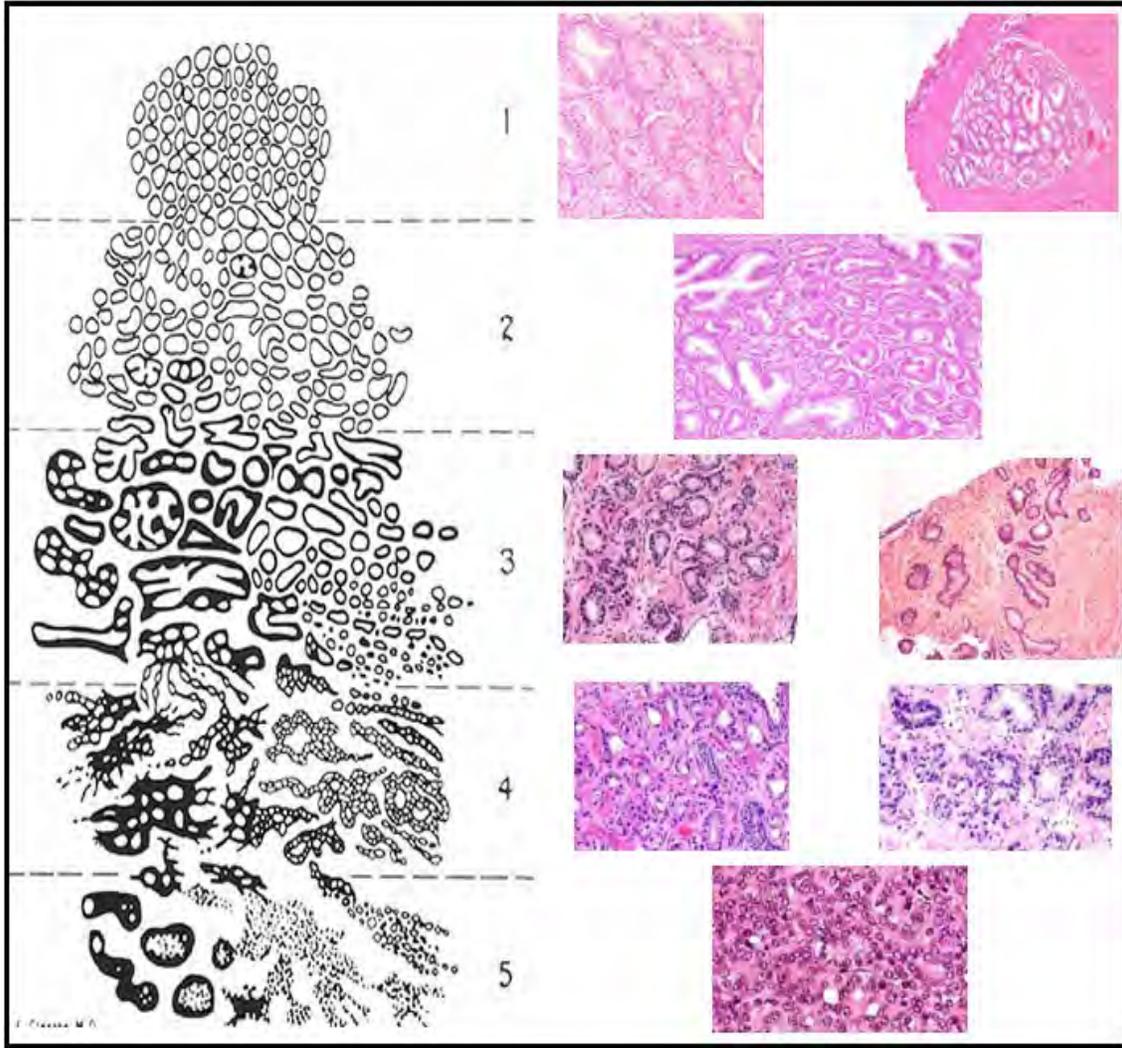
Figure 1.9: Gleason grading of prostate cancer, from least aggressive (grade 1 or normal) to most aggressive (grade 5) [14].

process includes three steps: (i) detecting cancer regions in the tissue, (ii) assigning a Gleason grade to each region and (iii) assigning a Gleason sum to the entire tissue. These steps are described as follows.

- To detect cancer regions in the tissue in step (i), the pathologist examines the tissue image at different magnifications. He usually starts with a low magnification (1× or 2×) to localize the regions of interest (regions with small glands or abnormal tissue structures) in the image. Next, he examines the image at 5× or 10× magnification to
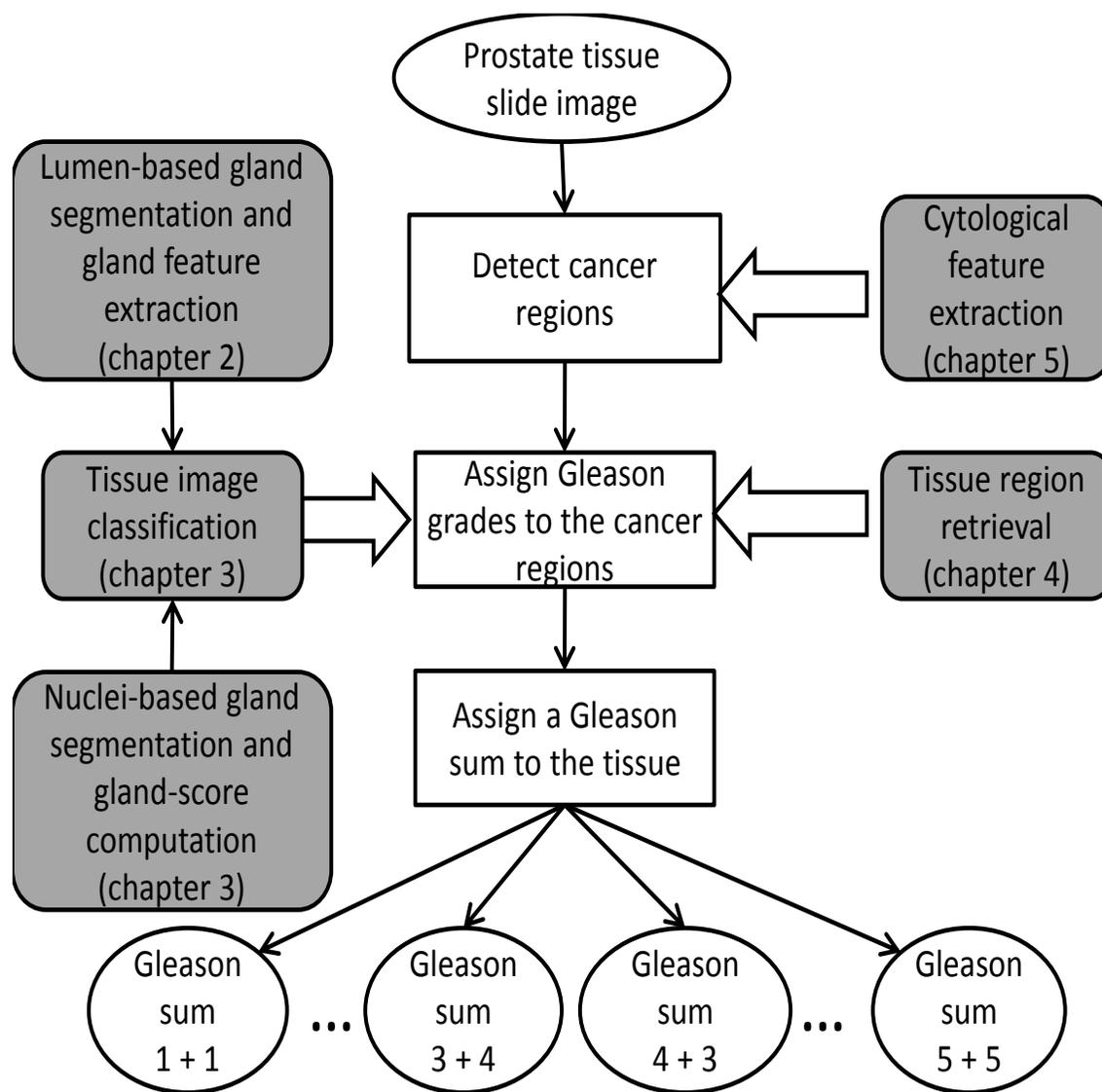
Figure 1.10: Histopathological grading process, using Gleason grading method, performed by a pathologist. The gray boxes indicate the modules designed and developed in the thesis, which can assist pathologists in this grading process.

analyze the gland structures and the distribution of glands in the tissue. For regions where gland structures do not provide sufficient evidence, the pathologist needs to magnify the image to $20\times$ or $40\times$ to examine the cytological information in the tissue, e.g. nuclei shape and size, the amount of cytoplasm, the presence of nucleoli inside the nuclei, etc.

- In step (ii), using the Gleason grading method described in [14, 35], the pathologist assigns a Gleason grade to each detected region obtained from step (i) based on the histological tissue pattern present in the region. This Gleason grading method defines five Gleason grades (from 1 to 5) corresponding to five different Gleason patterns (Figure 1.9). The Gleason grade 1 pattern is the least aggressive and the Gleason grade 5 pattern is the most aggressive. A low grade pattern (grade 1 or 2) is very similar to the normal pattern while a high grade pattern (grade 4 or 5) is very different from the normal pattern. Glands in grade 1 and grade 2 patterns have well-defined structures and are closely packed. Recent studies have recommended that Gleason grades 1 and 2 should rarely be used due to their uncommonness in the tissue [36, 37]. The grade 3 pattern, which is the most common pattern, usually has small glands infiltrating (invading) to the muscle (stroma). Glands in grade 3 pattern still form individual units and stay separated from each other. However, in grade 4 pattern, glands tend to fuse with nearby glands, so that the gland structures become ill-defined. As a result, individual gland units are not easy to identify in this pattern. Finally, grade 5 pattern is characterized by the total loss of gland structures, so that only sheets of cells are visible in this tissue pattern.

- Since a tissue may contain different Gleason patterns, the pathologist reports a *Gleason sum*[3] for the whole tissue (step (iii)). The Gleason sum is the sum of the grades of the most and the second most dominant Gleason patterns in the tissue. As a result, the

---

[3]The Gleason sum is also known as the Gleason score in some of the studies.

Gleason sum ranges from 2 (1 + 1) to 10 (5 + 5). However, a Gleason sum of $7 = 4 + 3$ is worse than a situation where the same Gleason sum is obtained as $7 = 3 + 4$. It is, thus, better to specify the breakdown of the Gleason sum into two numbers rather than simply reporting the total.

Although the biopsy only represents a portion of the prostate, the Gleason sum of the biopsy is still one of the strongest factors for prognosis. The Gleason sum can help physicians to determine the appropriate treatment. It is recommended that a biopsy be examined by more than one pathologist [15].

In this thesis, similar to other studies in computer-aided diagnosis of prostate cancer, we aim at providing tools to assist pathologists in performing the first two steps mentioned above, i.e. finding cancer regions and assigning Gleason grades to these regions (Figure 1.10), which are the fundamental tasks in the Gleason grading method. By using these tools, the pathologist can determine the Gleason sum of the tissue. More specifically, we consider the three most common types of regions in the tissue: normal regions (non-malignant glandular regions), Gleason grade 3 and Gleason grade 4 regions[4].

1.3   Challenges in Processing Prostate Tissue Slide Images

We summarize the major challenges in processing and interpreting prostate tissue slide images:

1. The first challenge relates to the large image size, as discussed above. We address this problem by processing the images at a lower magnification. Our methods for gland segmentation, gland classification, and tissue image classification obtain good results for images at 5× magnification, which are comparable to results for images at 20× magnification.

---

[4]Pathologists do not consider grade 1 and grade 2 when annotating the tissue images in our database. Moreover, grade 5 regions are only observed in a few tissue areas in the database. Hence, our primary interest here is in normal, grade 3 and grade 4 regions.

2. The second challenge is the large variations, both in shape and appearance, in the glandular structures, which makes model-based approaches such as Active Shape Model [38] or Active Appearance Model [39] inapplicable. Our proposed gland segmentation method is a model-free approach which is able to capture important structures of the glands regardless of their variations in size and shape.

3. The third challenge is the large color variations in the tissue structures in the image. This can be due to several reasons such as: (i) differences in staining (tissue is either overstained or understained), (ii) long term storage which makes the stain fade overtime, (iii) different types of tissue scanners. We address these variations as follows:

   a) We use the *Lab* color space to represent the image pixels. *Lab* color space, which separates the luminance and chrominance of the color, is suitable to describe the color information of the tissue components. For example, due to the blue color, nuclei appear most salient in the b channel of the *Lab* space (chapter 5). Moreover, each color channel L, a, b is normalized to $[0,1]$ to account for the difference in color intensity among images.

   b) We apply an adaptive binarization method (chapter 5) to segment nuclei. The threshold used to binarize the grayscale image is adaptively chosen for each image, instead of being set to a fixed value.

1.4   Related Work

The literature on prostate histopathological grading can be divided into two main categories, namely tissue image classification and cancer region detection. However, most of the studies address the tissue image classification problem. In this problem, the aim is to classify a tissue image into one of the two classes: normal and cancer, or into one of the multiple classes: normal, grade 3, 4 and 5 [40–44]. Only a few author have addressed the cancer detection problem, i.e., given a tissue image which includes both cancer regions and normal
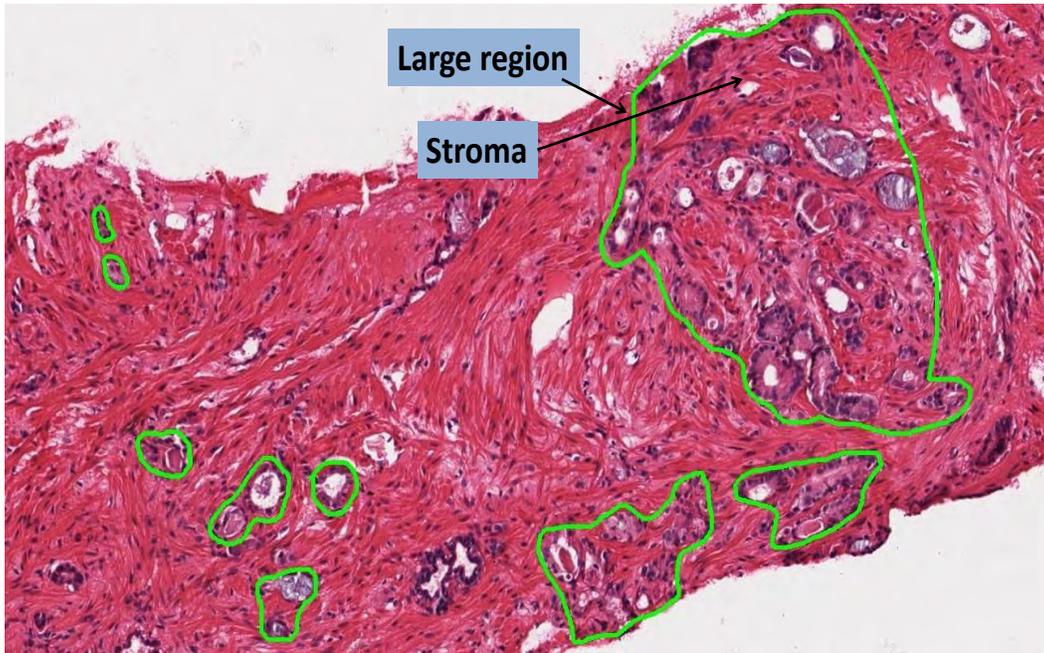
Figure 1.11: Annotation of cancer regions (green contours) by a pathologist. The annotation is a subjective process because the pathologist may circumscribe a large region (on the right) containing several cancer glands or several small regions (the remaining regions), each containing only a few glands. The image size is $600 \times 1{,}200$ pixels, corresponding to a 1.2×2.4 mm^2 tissue region digitized at $5\times$ magnification.

regions, automatically detect the cancer regions (regions of interest) in the image [3, 45, 46]. A major challenge in the cancer detection problem is that the ground truth data, namely the annotated cancer regions marked by pathologists, is not sufficient to accurately evaluate an automated cancer detection method (this was previously mentioned in [45]). The ground truth, created as a hand drawn contour by pathologists not only contains cancer glands but also contains stroma (which is really the background area). Furthermore, there are no standardized rules for the contour drawing, e.g., it may contain multiple glands or just a few glands, which is a subjective choice of the pathologist, as shown in Figure 1.11.

To address the two problems mentioned above, the published studies have either used a segmentation-based approach (gland segmentation followed by gland feature extraction) or used a segmentation-free approach (directly extracting texture features from the image

itself). We will discuss these methods in more detail in the subsequent chapters.

1.5   Contributions

This thesis makes the following contributions to CAD systems for histopathological grading of prostate cancer (Figure 1.10):

1. A lumen-based gland segmentation method, termed nuclei-lumen-association (NLA), that considers lumen as the central component of the gland, and find nuclei associated with the lumen. The NLA method leads to better gland classification results than the level set method, and comparable gland classification results to the manual gland segmentation method.

2. The gland features that capture rich information about the gland, including structural and contextual information. We show that these proposed gland features are (i) better than gland features used in the published studies in solving the gland classification problem and (ii) better than texture features used in the published studies in solving the tissue image classification problem.

3. A nuclei-based gland segmentation method that models the relationship between nuclei and lumina in the image by a nuclei-lumina-graph. The edges in the graph correspond to the links between nuclei and nuclei and between nuclei and lumina. The recursive normalized cut method is applied to partition the graph into different components, each of which corresponds to a gland. The nuclei-based method overcomes the major limitation of the conventional lumen-based methods: it is able to detect glands without lumen and glands with multiple lumina.

4. A method to compute the gland-score of a segmented region that measures how similar the segmented region is to a gland unit. The gland-scores are combined with the

structural-contextual features to improve the grade 3 vs grade 4 tissue image classification result.

5. Both the lumen-based and nuclei-based methods are evaluated on both low magnification ($5\times$) and high magnification images ($20\times$) to evaluate the effect of image magnification on the performance of these methods.

6. A gland-based method to compute the similarity between two tissue regions. This method can be used to search for regions similar to a region of interest (ROI) in the annotated tissue slides. The retrieved regions can serve as the references that a medical student or a technician can rely upon when grading the ROI.

7. A method to extract an important cytological feature in the prostate tissue images, i.e. the presence of nuclei with prominent nucleoli. To our knowledge, this cytological feature has not yet been exploited in any CAD system. We also demonstrate the usefulness of this cytological feature by applying it in a prostate cancer detection framework.

1.6 Database Used in the Thesis

The prostate tissue image database used in our empirical studies contains 39 tissue slides taken from 29 patients. The database was provided by Ventana Medical Systems, Inc. [23]. The tissue slides, with an average size of 5.6×9.2 mm^2, were stained using the H&E method. These slides were digitized at $20\times$ magnification, yielding images of approximately $11,150 \times 18,300$ pixels (the image resolution is approximately 0.5μm per pixel). In each tissue slide image, a pathologist annotates the Gleason grade 3 and Gleason grade 4 areas (if they are present). An example of a tissue slide image in the database with pathologist's annotations is shown in Figure 1.12. From the tissue slide images in this database, which we refer to as the *main database*, we obtain a sub-database for each of the research problems that we address in the subsequent chapters. See Table 1.3. The magnification of the tissue slide images (or

tissue image regions) in each sub-database can be 20× or 5×, which is chosen by taking into account the computational cost, the goal, and the requirement of the corresponding problem. For the gland segmentation and classification problems (chapter 2), we use tissue image regions at 5× magnification because using images at this magnification is sufficient to separate the three classes of interest: artifact, normal, and cancer glands (as mentioned in chapter 2, section 2.5.2). For the tissue image region classification problem (chapter 3), we use tissue image regions at both 20× and 5× magnification since we want to analyze the effect of image magnification on the classification accuracy. Due to the very large size of the entire tissue slide images at 20× (11,150 × 18,300 pixels), performing tissue image region retrieval (chapter 4) on these entire images at this magnification will be very computationally expensive. Moreover, we observe from the tissue image region classification results that the gland features computed from 5× images have almost the same discriminative information as those computed from 20× images. Hence, we only use 5× images when addressing the tissue image region retrieval problem. Finally, to compute the cytological features (chapter 5), we need to use images at 20× magnification because the cytological information in the tissue can only be discerned in high magnification images.

1.7   Thesis Organization

In chapter 2 we present the methods for gland segmentation (lumen-based method) and gland classification. In chapter 3 we discuss how to utilize these methods to solve the tissue image classification problem. We also introduce a new gland segmentation method (nuclei-based method), and introduce a method to compute the gland-scores of the segmented regions to improve the performance of Gleason grade 3 vs grade 4 tissue image classification. In chapter 4, we address the tissue region retrieval problem. In chapter 5, we present a method to exploit the cytological feature of the prostate tissue. Finally, chapter 6 summarizes the thesis and identifies some problems for future work.

Problem	Chapter	Sub-database
Gland segmentation and classification	2	48 tissue image regions at 5× magnification (average size is 900 × 1,500 pixels). These image regions are sampled from the tissue slide images of the main database in such a way that they contain glands with large structural variations. We select 525 artifacts, 931 normal glands, and 1,375 cancer glands from these image regions
Tissue image region classification	3	317 tissue image regions at both 20× and 5× magnification (average size at 20× magnification is 1,400 × 1,380 pixels). These image regions are sampled from the tissue slide images of the main database in such a way that each region corresponds to one Gleason grade. Among them, there are 113 normal, 134 Gleason grade 3, and 70 Gleason grade 4 regions
Tissue image region retrieval	4	All 39 tissue slide images in the main database at 5× magnification (average size is 2,800 × 4,500 pixels)
Cytological feature extraction	5	17 (out of 39) tissue slide images at 20× magnification (average size is 3,400 × 10,000 pixels), in which the cytological features are salient

Table 1.3: The sub-databases extracted from the main database that are used in different research problems in the thesis.
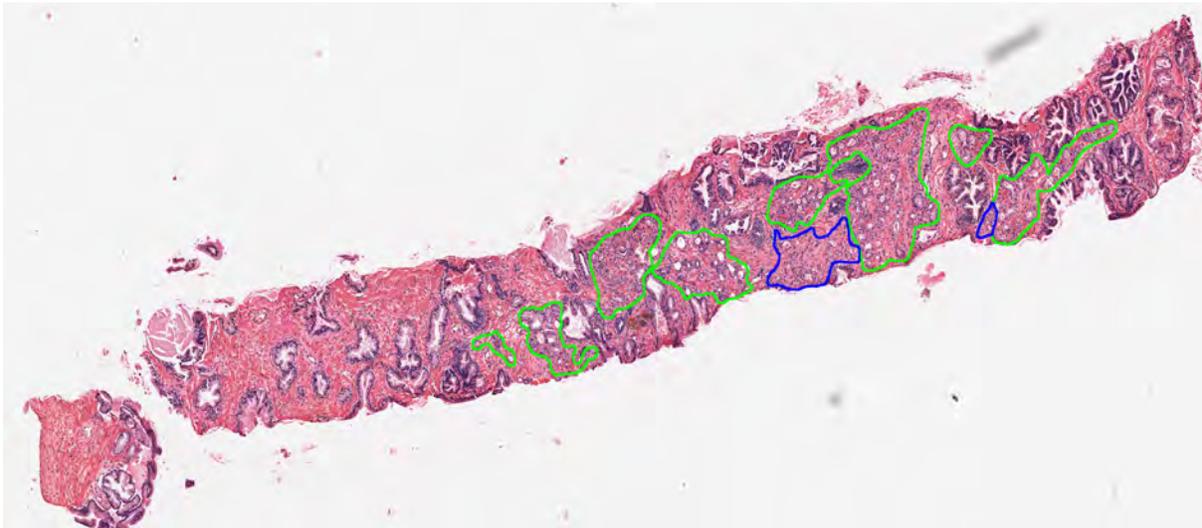
Figure 1.12: A tissue slide image in the main database. In this image, the pathologist annotates the Gleason grade 3 areas by the green contours and annotates the Gleason grade 4 areas by the blue contours. The size of the image is 1,160 × 3,100 pixels, corresponding to a 2.3 × 6.2 mm^2 tissue slide digitized at 5× magnification.

Chapter 2

GLAND SEGMENTATION AND CLASSIFICATION

2.1 Introduction

According to the Gleason grading process mentioned in chapter 1, Figure 1.10, the pathologist needs to first find cancer regions in the tissue slide (Figure 2.1). In order to perform this task, the pathologist relies on: (i) structural information; glands in a cancer region (which we call cancer glands) appear to have structural properties (such as nuclei abundance, lumen size) different from glands in a normal region (which we call normal glands) and (ii) contextual information; cancer glands typically cluster together to form a group and are of similar shape and size[1], while the shape and size of normal glands vary widely. These two sources of information can be observed in Figure 2.1. Consequently, a reasonable approach to develop an automatic system to assist a pathologist in finding cancer regions is to first detect individual cancer glands. This can be accomplished by segmenting glands, examining their structural and contextual properties and finally classifying them as cancer glands or normal glands. Since artifacts, which are created by broken tissue areas (Figure 2.2), are commonly present in the tissue images, we need to distinguish them from the true glands. As a result, we address a three-class classification problem: artifacts, normal glands, and cancer glands.

Challenges in Gland Segmentation and Gland Classification

The challenges of the problem are the following (Figure 2.3):

- Glands have large variations in shape and size, i.e. the shape can be circular, ellipsoid, or can be very complex, while the size can range from small (2,500 pixels) to large (2,900,000 pixels).

---

[1]It was also mentioned in [47] that cancer glands tend to appear close to other cancer glands, which is a domain knowledge supporting this contextual information
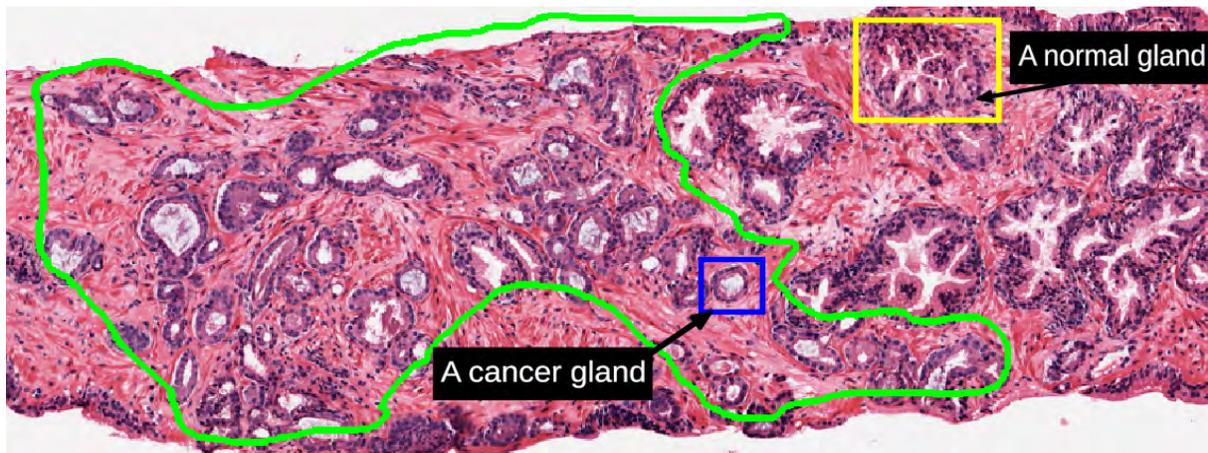
Figure 2.1: A tissue image showing the cancer glands in a cancer region annotated by a pathologist (green contour); normal glands are present in the region outside the green contour. The image size is 370×900 pixels, corresponding to a 0.74×1.8 mm^2 tissue region digitized at $5\times$ magnification.
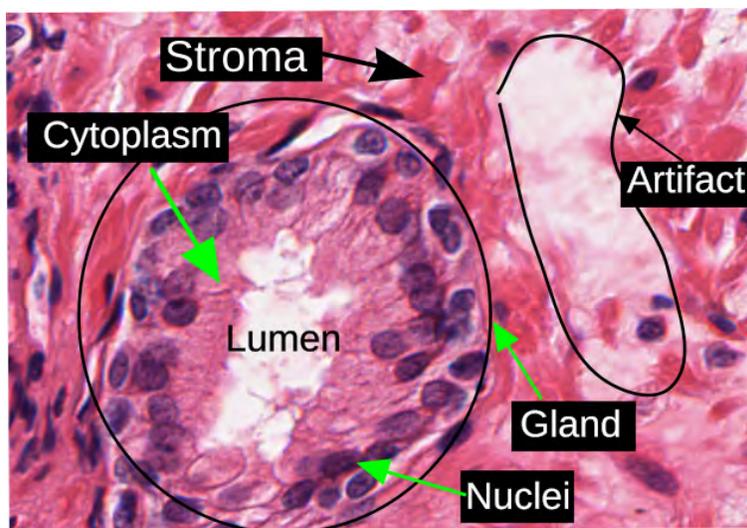
Figure 2.2: A gland with basic components (nuclei, cytoplasm and lumen) and an artifact. Stroma can be considered the tissue background. In this H&E stained image, nucleus mostly appears blue, cytoplasm mostly appears purple, lumen and artifact mostly appear white and stroma mostly appears pink. The image size is 230×330 pixels, corresponding to a 0.12×0.16 mm^2 tissue region digitized at $20\times$ magnification.

24

- The color intensity of different components in the tissue may change remarkably from slide to slide. This is caused by the variations in the staining procedure, the thickness of the tissue, the amount of stain used, or the duration of tissue storage.

We address the color variation problem by using an unsupervised method (k-means clustering) to identify the tissue components, rather than applying a supervised method like in [2], which relied on a fixed training set of pixels for each tissue component. Moreover, we use the *Lab* color space, which separates the chrominance and luminance of the color, to represent the image pixels. Given the large variation in gland size and shape, model-based methods such as Active Shape Model [38] or Active Appearance Model [39] are not applicable. So, we developed a model-free method, which is based on the structures of the glands, for gland segmentation.
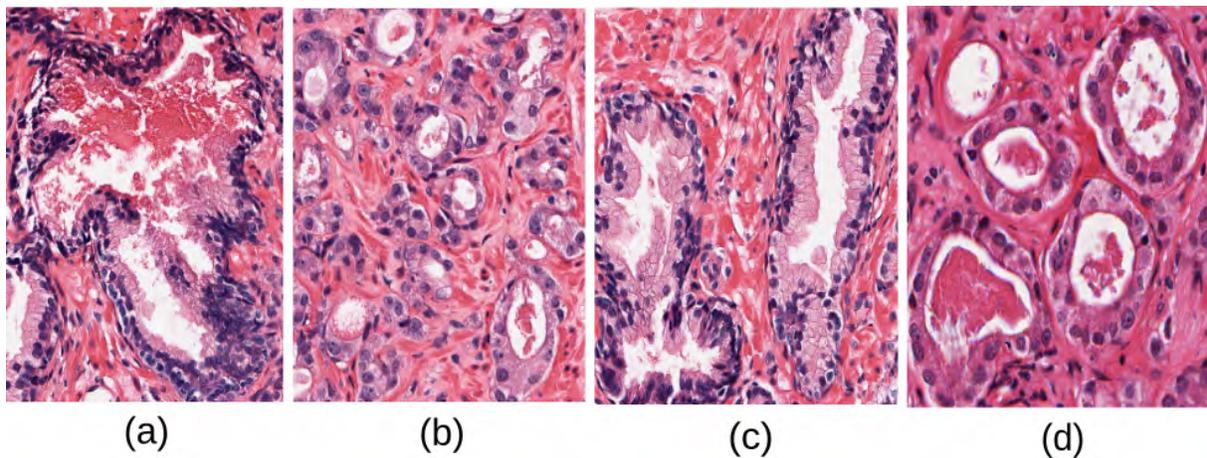
Figure 2.3: Challenges in gland segmentation and classification. (a)-(b) Variations in gland structure; a large gland with irregular shape (a) and small glands with circular shape (b). (c)-(d) Variations in color intensity; in (c) the cytoplasm color is very different from stroma but in (d) the cytoplasm color is similar to stroma (d). Images in (a) and (c) are obtained from normal tissue regions while images in (b) and (d) are obtained from cancer tissue regions. The average size of the images is 200×180 pixels, corresponding to a 0.4×0.36 mm^2 tissue region digitized at $5\times$ magnification.

2.2 Related Work

2.2.1 Gland Segmentation

The gland segmentation problem has been studied in the literature. Two popular methods for gland segmentation are level set [2, 48, 49] and region growing [3, 50]. In the level set method, basically, a closed curve is evolved and subjected to some constraints in such a way that it can detect an object of interest in the image. For example, one can initialize the curve inside the object of interest and evolve the curve toward the object boundary. For a formal definition [51], a level set function is defined as $\phi(t, x, y)$, where t denotes the time and x, y denote the 2D coordinates of the image pixels. Next, the zero level curve is defined as $C = \{(x, y)|\phi(t, x, y) = 0\}$, and is used to detect the object of interest in the image. The zero level curve C is implicitly evolved by using the level set equation:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0 \tag{2.1}$$

where F is the speed function which depends on the image. For example, F can be defined based on the gradient of the image so that the curve moves toward the object boundary. In practice, the level set evolution is usually solved by a variational method, i.e. minimizing an energy functional which is defined on the level set function. In general, the total energy function of a level set includes two terms:

$$\mathcal{E}(\phi) = \mu \mathcal{P}(\phi) + \mathcal{E}_{ext}(\phi). \tag{2.2}$$

The first term $\mathcal{P}(\phi)$ denotes the internal energy (which maintains the stability of the curve), while the second term $\mathcal{E}_{ext}(\phi)$ denotes the external energy (which drives the curve to the region of interest (ROI) in the image). By simultaneously minimizing both the internal and external energy, the zero level curve is evolved to the ROI while still maintaining its stable form. Figure 2.4 illustrates the use of the level set method for gland segmentation as presented in [2]. In this case, the ROIs are the nuclei on the gland boundary.

For the region growing method, we can analyze the algorithm used in [3] as an example.

26
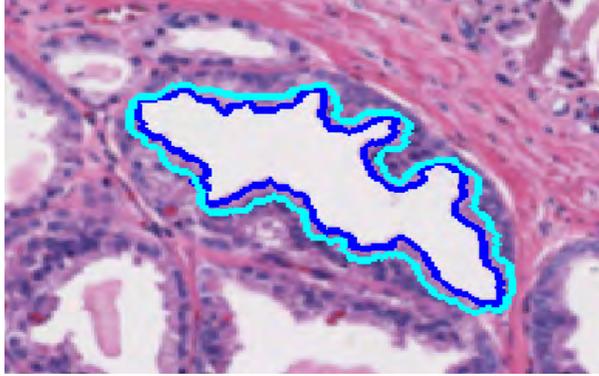
Figure 2.4: Use of the level set method for gland segmentation as presented in [2]. The blue contour denotes the boundary of the lumen where the level set curve is initialized, while the cyan contour denotes the level set curve after the evolution.

In this algorithm, the seed points, which are considered centers of glands, are first selected as the high intensity pixels in the luminance channel of the image. The region growing algorithm is performed in an iterative manner at every seed. Basically, at every step, (i) the pixel with the highest intensity in the current boundary of the growing region is used to expand the region and (ii) the boundary strength of the growing region at this step is computed using the current boundary. The optimal region is selected as the region with the largest boundary strength during the expansion. Figure 2.5 illustrates the use of the region growing method for gland segmentation as presented in [3].

2.2.2 Gland Feature Extraction and Classification

In order to detect cancer regions in a tissue image, Monaco et al. [3] first used gland size to classify glands in the image into normal or cancer glands before applying the probabilistic pairwise Markov model (PPMM) to update the gland labels. Adjacent cancer glands were then grouped together to form cancer regions.

The PPMM, which is a variation of the Markov Random Field (MRF) model, was used to capture contextual information about the glands. This model is briefly described as follows. Given an image with n glands, let $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ denote the gland labels, and $\mathbf{y} =$

27
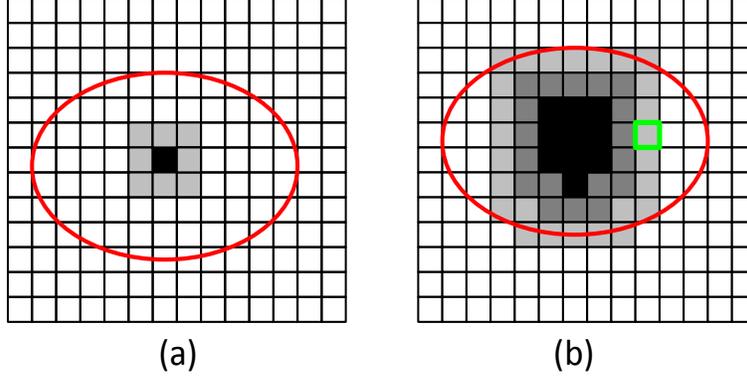
Figure 2.5: Use of the region growing method for gland segmentation as presented in [3]. The red contours in (a) and (b) denote the lumen boundary. The black square in (a) denotes the point with highest intensity being chosen as the seed, while the gray squares in (a) denote the boundary of the seed. The non-white region in (b) denotes the region growing from the seed in (a), where the green rectangle denotes the point with the highest intensity on the current boundary.

$\{y_1, y_2, ..., y_n\}$ denote the gland features. The gland labeling (gland classification) procedure is equivalent to a maximum a posteriori (MAP) estimation procedure, i.e., choosing the gland labels $\mathbf{x}$ in such a way that $P(\mathbf{x}|\mathbf{y})$ is maximized:

$$P(\mathbf{x}|\mathbf{y}) = P(x_g|\mathbf{x}_{-g}, \mathbf{y})P(\mathbf{x}_{-g}|\mathbf{y}) = P(x_g|\mathbf{x}_{\eta g}, y_g)P(\mathbf{x}_{-g}|\mathbf{y}), \qquad (2.3)$$

where x_g is the label of the gland g, and $\mathbf{x}_{\eta g}$ denotes the labels of neighboring glands of g (the neighboring glands of g were defined as glands within a distance $R = 0.9$ mm from g). This maximization problem can be solved by the iterated conditional modes (ICM) technique [52]: each gland g is iteratively visited and a label x_g is chosen to maximize $P(x_g|\mathbf{x}_{\eta g}, y_g)$. Moreover, $P(x_g|\mathbf{x}_{\eta g}, y_g) \sim p(y_g|x_g)P(x_g|\mathbf{x}_{\eta g})$, where the density $p(y_g|x_g)$ and the probability $P(x_g|\mathbf{x}_{\eta g})$ can be learned from a training set of labeled glands. In general, the PPMM method is used to iteratively update gland labels after the initial gland labeling is performed using the gland features alone.

Unlike Monaco et al. [3], Peng et al. [50] and Naik et al. [2] used gland features to classify the tissue image rather than explicitly classifying the individual glands. While Peng et al. [50] only used gland size to classify the tissue images into normal and cancer images, Naik et al. [2] used size and shape of both gland and lumen to classify a tissue image into normal,

28
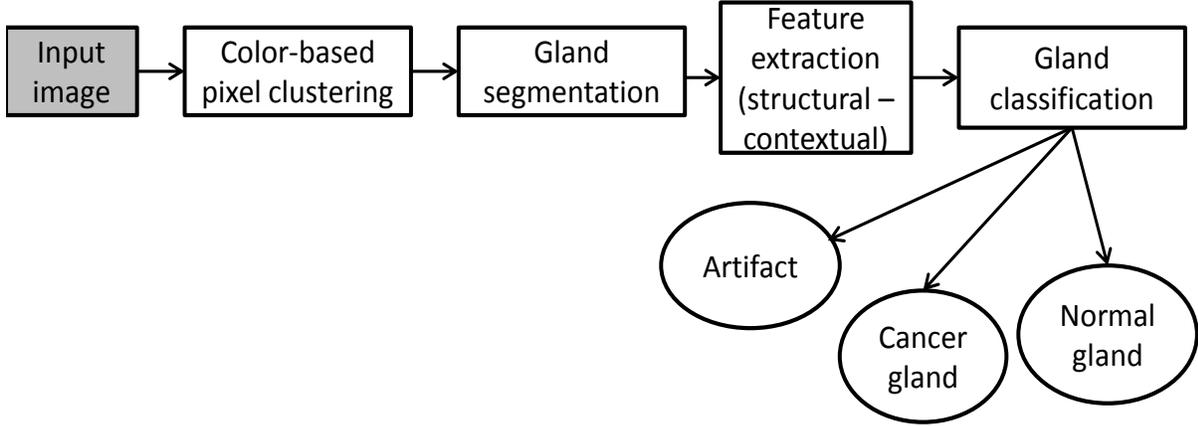
Figure 2.6: Proposed method for gland segmentation and classification.

Gleason grade 3 or Gleason grade 4. By denoting the area, perimeter and boundary contour of the lumen as A, P, and B, respectively, they computed the following shape features for the lumen:

1. Area overlap ratio feature: Ratio of A to the smallest circular area that encloses B.

2. Distance features: Given the set of distances d from the center of B to the points on B, these features include maximum(d)/mean(d), and the variance of d.

3. Perimeter feature: P'/P, where P' denotes the estimated perimeter of the lumen or gland (obtained by interpolating points equally sampled on B).

4. Compactness feature: P^2/A.

5. Smoothness feature: For every point p_i in B, compute $S_{p_i} = |d(p_i, p_g)(d(p_{i-1}, p_g) + d(p_{i+1}, p_g))/2|$, where p_g is the centroid of the lumen or gland. The smoothness feature is defined as $\sum_i S_{p_i}$.

The same shape features described above are also computed for the gland. A summary of these studies, along with the proposed method, is given in Table 2.1.

Study	Segmentation algorithm	Gland features	Use of context	Objective	Limitation
Monaco et al. [3]	Region growing	Gland size	PPMM method	Detect cancer regions	Only one feature, gland area, was used
Peng et al. [50]	Region growing	Gland size	No	Classify tissue images into normal and cancer images	Only gland area was used; contextual information was not considered
Naik et al. [2]	Level set	Shape features of lumen and gland	No	Classify tissue images into normal, Gleason grade 3 and grade 4 images	No features about nuclei; contextual information was not considered
Proposed method	Nuclei-lumen-association (NLA)	Structural features	Contextual features	Classify glands into artifacts, normal and cancer glands	The NLA method may not obtain good segmentation results for glands with weak nuclei on the boundary

Table 2.1: Gland segmentation and gland feature extraction methods for prostate tissue images reported in the literature. Note that each study used a different database since there is no public domain database available.

2.3 Proposed Gland Segmentation Method

2.3.1 *Lab* Color Space

Similar to [3], we use the *Lab* color space for image pixel representation. The *Lab* color space separates chrominance and luminance of the color, thus, making it suitable to describe the color information of the tissue components. The *Lab* (also known as L*a*b* or CIELAB) color space [53,54] is specified by the CIE (International Commission on Illumination) to separate the lightness of the color (L component) from the spectral properties of color (negative values of a component indicate green and positive values of a indicate red, while negative values of b component indicate blue and positive values of b indicate yellow). To convert

from RGB space to Lab space, we first convert RGB to XYZ space by:

$$
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4900 & 0.3100 & 0.2000 \\ 0.1769 & 0.8124 & 0.0107 \\ 0.0000 & 0.0099 & 0.9901 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.4}
$$

The elements of the transformation matrix were derived in [55]. Next, the XYZ space is converted to Lab space by:

$$
L = 116 f(Y/Y_n) - 16 \tag{2.5}
$$

$$
a = 500[f(X/X_n) - f(Y/Y_n)] \tag{2.6}
$$

$$
b = 200[f(Y/Y_n) - f(Z/Z_n)] \tag{2.7}
$$

where:

$$
f(t) = \begin{cases} t^{1/3} & \text{if } t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \tag{2.8}
$$

and X_n, Y_n, Z_n are the white point tristimulus values in XYZ [54].

2.3.2 Gland Structure

A gland consists of epithelial nuclei, epithelial cytoplasm and lumen (Figure 2.2, [2], [56]). In between the glands is the stroma which can be considered as the background. Artifacts appear as non-lumen white regions in the tissue. We perform gland segmentation by employing the following two steps.

2.3.3 Tissue Component Identification

To facilitate segmentation, we first identify the basic components in the tissue by using a color-based clustering procedure. By utilizing the differences in color of the four tissue components in an H&E image [57] (stroma (S) color is mostly pink, lumen (Lu) is mostly white, nucleus (N) is mostly blue and cytoplasm (C) is mostly purple), we perform clustering

in the *Lab* color space by using the k-means algorithm ($k = 4$). For computational efficiency, in each image, we run the k-means algorithm on 10,000 randomly selected pixels to find 4 clusters. By using the L and a values of the cluster centers, denoted by L^0 and a^0, respectively, we are able to match the four clusters with the four tissue components (nuclei, stroma, cytoplasm, and lumen) by the two following rules (see Figure 2.8b for an illustration):

1. $L^0_N < L^0_S, L^0_C < L^0_{Lu}$: Nuclei cluster has the lowest L value (since nuclei have the darkest color), while lumen cluster has the highest L value (since lumen has the brightest color).

2. $a^0_S > a^0_C$: Stroma cluster has a higher a value than cytoplasm cluster since the color of stroma is more red than cytoplasm.

Next, each pixel in the image is assigned to the tissue component corresponding to the nearest cluster center (Figure 2.8c). We apply a connected component algorithm [58] on nuclei pixels and lumen pixels to generate nuclei objects and lumen objects, respectively, which are later used for segmentation. Although the list of lumen objects also contains artifacts, we cannot discard them easily by a simple size thresholding operation because these artifacts can be as large as the lumen.

2.3.4 Nuclei-Lumen Association

The gland segmentation algorithm starts with lumen as the first component of a gland and then searches for the nuclei for that gland. The proposed algorithm (Figure 2.7), which is referred to as nuclei-lumen association (NLA) algorithm, associates appropriate nuclei with each lumen to create a gland segment. Nuclei are searched along a direction normal to the lumen boundary contour. The algorithm consists of three steps:

1. Given n points on the lumen boundary, we sample $n/3$ points uniformly at equal interval, called *lumen points*. The number of points is selected by considering the

trade-off between a sparse (for computational efficiency) and dense set (for adequate search coverage).

2. A search region (Ω_S) of a conical shape, centered at each lumen point, is expanded to find nuclei on the gland boundary. When nuclei pixels are found within Ω_S, we use a circular mask to select the nuclei region to join the gland boundary. This step is repeated for all lumen points to find the complete gland boundary.

3. A pruning procedure, based on the median absolute deviation (MAD), is applied to remove outlier nuclei and generate a smoother segmentation boundary[2].

Besides gland segments, the proposed algorithm also produces a set of points located at the detected nuclei, referred to as the nuclei point set. The nuclei point set and lumen point set are used for gland feature extraction. A detailed description of the algorithm is given in Algorithm 2.1.

Since artifacts are included in the list of lumen objects, some non-gland segments created by artifacts are also present in the segmentation result (Figure 2.9a). However, instead of detecting them at this step, we identify them in the classification procedure, which is more reliable since it uses a larger number of features.

2.4 Gland Feature Extraction

Given that artifacts are also present at the output of the segmentation stage, we classify a gland segment into an artifact, a normal gland or a cancer gland. The main differences in structures of the three classes of glands (artifact, normal gland and cancer gland) are as follows:

---

[2]We define $\mathrm{MAD} = \mathrm{median}_i(|d_i - \mathrm{median}_j(d_j)|)$, where d_i and d_j are the distances between a lumen point and a nuclei point. A nuclei point k with $d_k > [3\sigma + mean(d_i)]$, where $\sigma = 1.48\mathrm{MAD}$, is considered an outlier and discarded.

Algorithm 2.1 Nuclei-lumen association (NLA) algorithm

Input: Lumen object Lu, m nuclei objects $\{N_i\}_{i=1}^m$. Image intensify function f. Parameters α_0, η, r_C, r_m

Output: Gland segment GS, lumen point set $\{L_p\}$, nuclei point set $\{N_p\}$

$\quad GS \leftarrow Lu$ (gland is initialized as lumen)

2: Select $\{L_p\}$ by sampling at equal intervals from the boundary points of Lu

$\quad \{L_p\} = \{p_j\}_{j=1}^n$; $\{N_p\} \leftarrow \{q_j = (0,0)\}_{j=1}^n$

4: **for** $j = 1$ to n **do**

$\quad\quad r \leftarrow 0$ ((r, α_0) determines the area of the search region at point j, Ω_S^j)

6: $\quad$ Let $p_j = (p_{jx}, p_{jy})$. Compute $\nabla f_j = \left(\frac{\delta f}{\delta p_{jx}}, \frac{\delta f}{\delta p_{jy}} \right)$

$\quad\quad flag \leftarrow$ FALSE

8: $\quad$ **while** $flag =$ FALSE **do**

$\quad\quad\quad r \leftarrow r + \eta$ (expand Ω_S^j)

10: $\quad\quad \Omega_S^j = \{(x,y) | \|(x - p_{jx}, y - p_{jy})\|_2 < r$ and $\|\angle(x - x_j, y - y_j) - \angle(\nabla f_j)\| < \alpha_0\}$

$\quad\quad$ Let $\Lambda_N = \{N_i, i = 1, \ldots, m$ such that $N_i \cap \Omega_S^j \neq \emptyset \}$ (the nuclei objects being found)

12: $\quad\quad$ **if** $\Lambda_N \neq \emptyset$ **then**

$\quad\quad\quad$ Let $x_0 = \text{mean}(x)$, $y_0 = \text{mean}(y)$, where $(x,y) \in \Omega_S^j \cap \Lambda_N$

14: $\quad\quad\quad \Omega_C = \{(x,y) | \|(x - x_0, y - y_0)\|_2 \leq r_C\}$ (circular mask)

$\quad\quad\quad GS \leftarrow GS \cup (\Omega_C \cap \Lambda_N)$ (take nuclei region as part of gland boundary)

16: $\quad\quad\quad q_j \leftarrow (x_0, y_0)$ (a nuclei point is found)

$\quad\quad\quad flag \leftarrow$ TRUE

18: $\quad\quad$ **end if**

$\quad\quad$ If Ω_j intersects any lumen object or exceeds a maximum size r_m, $flag \leftarrow$ TRUE

20: $\quad$ **end while**

$\quad$ **end for**

22: **Gland segment pruning:**

$\quad$ Remove p_j, q_j from $\{L_p\}, \{N_p\}$ where $q_j = (0,0)$

24: Let $d_j = \|(p_{jx} - q_{jx}, p_{jy} - q_{jy})\|_2 \ \forall p_j \in \{L_p\}$ and $q_j \in \{N_p\}$

$\quad MAD = median_i(\|d_i - median_j(d_j)\|)$

26: $\sigma \sim 1.48 MAD$

$\quad d_m \sim 3\sigma + \text{mean}(d)$

28: $\forall d_j$ such that $d_j > d_m$ (outliers), discard p_j, q_j

---
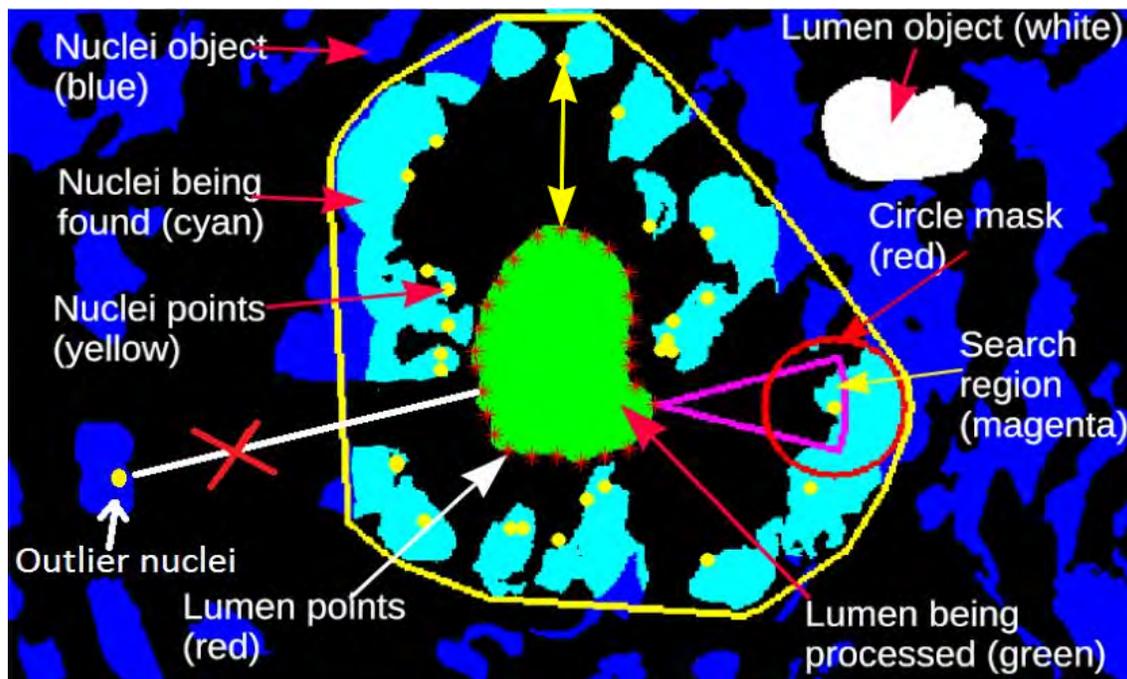
Figure 2.7: Nuclei-lumen association. The final segmentation result is depicted by a convex hull enclosing the detected nuclei.

- An artifact (Figure 2.9a) does not have cytoplasm surrounding the lumen and has very few associated nuclei.

- Nuclei on the boundary of a normal gland (Figure 2.9b) are more abundant and have darker blue color than a cancer gland (Figure 2.9c).

- Lumina of cancer glands commonly appear more circular and smaller in size than normal glands (Figure 2.1), while lumina of normal glands are more branchy and have large variation in shape.

Based on these differences, we extract the following sets of *structural features*, for a total of 19 features, for each gland:

1. **Set 1 (8 nuclei features)**: For each nuclei point (N_p), we compute the mean (μ) and standard deviation (σ) of the L, a, b color bands in the neighborhood of N_p (we denote this neighborhood by Ω_{N_p}). This color information describes the darkness of the blue
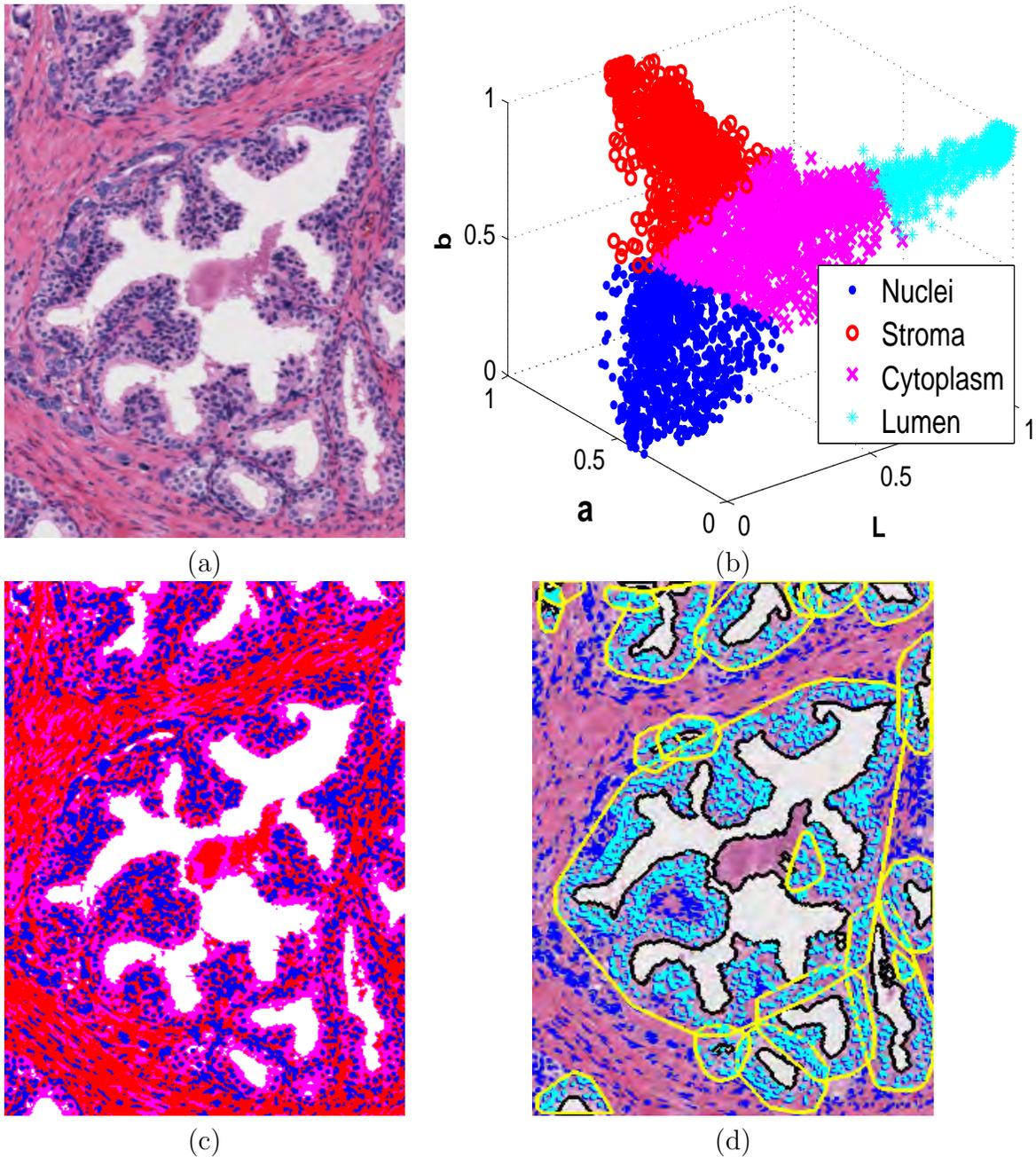
Figure 2.8: Proposed gland segmentation process. (a) Input tissue image of 390×270 pixels, corresponding to a 0.78×0.54 mm^2 tissue region digitized at $5\times$ magnification. (b) Clustering of the image pixels in the *Lab* color space. (c) Tissue component identification result where red denotes stroma, pink denotes cytoplasm, blue denotes nuclei, and white denotes lumen. (d) Nuclei-lumen association result (final gland segmentation result), where blue regions denote nuclei, black contours denote lumen, cyan regions denote nuclei associated with the lumen in the center, and yellow contours denote the segmented gland regions.
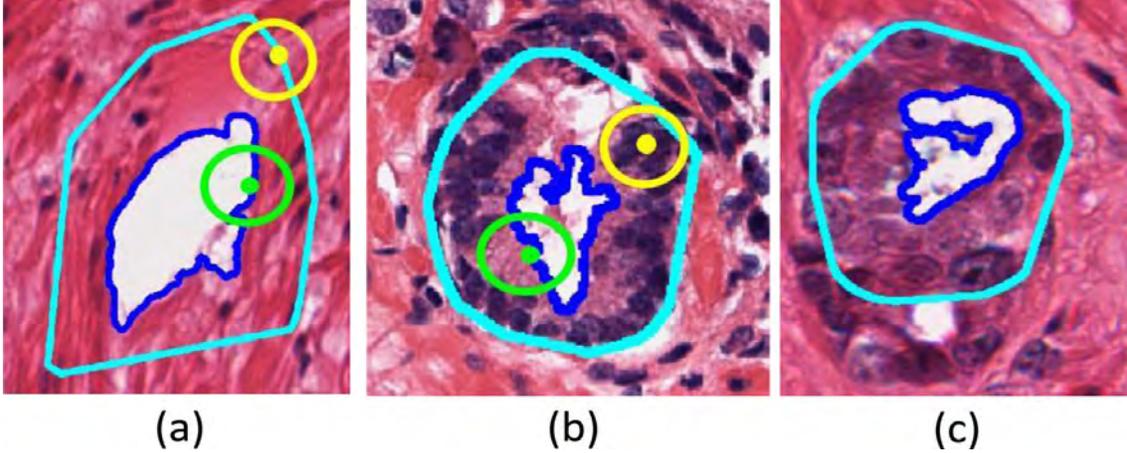
Figure 2.9: Segmentation results for three classes of interest: (a) Artifact, (b) normal gland and (c) cancer gland. Green and yellow circles in (a) and (b) denote the neighborhood of a lumen point and a nuclei point, respectively. The average size of the images is 180×160 pixels, corresponding to a $0.09 \times 0.08 \, \text{mm}^2$ tissue region digitized at $20\times$ magnification.

color in the nuclei. In addition, we compute μ and σ of the percentage of Ω_{N_p} that contains nuclei pixels, i.e. nuclei abundance on the gland boundary and its variation. Ω_{N_p} is a circular region centered at N_p (yellow circles in Figures 2.9a and 2.9b), and has a radius r_{N_p}. Since cancer glands usually have one nuclei layer on the boundary, while normal glands have more than one nuclei layer (mostly 2 nuclei layers), we choose $r_{N_p} = 10$ pixels, which corresponds to the 2-nuclei-layer thickness (the diameter of a nucleus is approximately 5 pixels in $5\times$ images). Hence, Ω_{N_p} is sufficient to capture most of the nuclei of a gland, while excluding nuclei of neighboring glands.

2. **Set 2 (6 cytoplasm features)**: For each lumen point (L_p), we compute mean and standard deviation of the L, a, b color bands in the neighborhood of L_p (we denote this neighborhood by Ω_{L_p}). Ω_{L_p} is a circular region (green circles in Figures 2.9a and 2.9b), which is centered at L_p and excludes lumen area. The radius of Ω_{L_p} is the distance between L_p and the corresponding N_p, which means Ω_{L_p} captures the region in between lumen and nuclei. This region mostly contains epithelial cytoplasm in a true gland segment (since cytoplasm is located between lumen and nuclei), and mostly contains stroma in an artifact segment (since artifacts are not glands, thus, are

37

not surrounded by epithelial cytoplasm). As a result, the color information in Ω_{L_p} is different for true glands and artifacts.

3. **Set 3 (3 lumen shape features)**: Area, solidity (ratio of the lumen area to its convex hull area) and circularity $((4\pi\text{area})/\text{perimeter}^2)$ of the lumen. For lumen that is more circular and compact (less branchy), the circularity and solidity are high.

4. **Set 4 (2 morphological features)**: Mean and standard deviation of the distance between a L_p and a N_p (yellow arrow in Figure 2.7). This distance is large for the artifact segments because artifacts are not surrounded by epithelial nuclei like the true glands.

As discussed at the beginning of the chapter, cancer glands tend to be located close to each other as a group while normal glands are distributed randomly. This motivates us to utilize contextual information in the classification. To explore contextual information, we first assign gland segments into groups (Figure 2.10) by using the connected component algorithm. Let $\{Lu^i\}_{i=1}^n$ denote the n lumen objects used to represent n gland segments, and let $(Lu_{x_o}^i, Lu_{y_o}^i)$ denote the centroid of Lu^i. A graph is built, where each node is a gland. If $\|(Lu_{x_o}^i - Lu_{x_o}^j, Lu_{y_o}^i - Lu_{y_o}^j)\| < t_d$, there is an edge connecting Lu^i and Lu^j. Each connected component is considered a group of glands. Once groups are formed, we compute the following 3 *contextual features* for each gland segment Lu^i (which belongs to group O):

1. **Neighborhood crowdedness**: $|O|$ or the number of elements in O.

2. **Shape similarity**: $\frac{1}{|O|}\sum_{j=1}^{|O|}\|LV_i - LV_j\|$, where LV denotes the 3-dimensional lumen shape feature vector described above.

3. **Size similarity**: $\frac{1}{|O|}\sum_{j=1}^{|O|}\frac{min(|Lu^i|,|Lu^j|)}{max(|Lu^i|,|Lu^j|)}$, where $|Lu|$ denotes the lumen size.

Finally, each gland segment is represented by a feature vector of dimensionality 22 (19 + 3). Table 2.2 summarizes the 22 gland features described in this section. We use an SVM
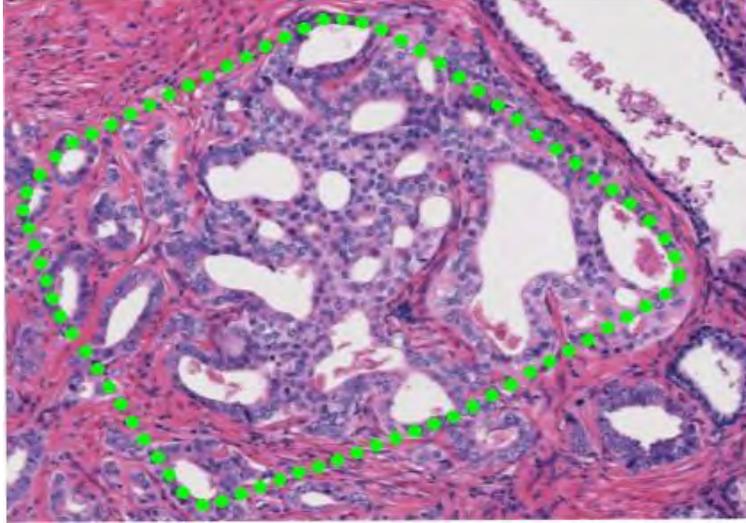
Figure 2.10: Glands are clustered into a group (dotted contour) to compute contextual features. The image size is 340×420 pixels, corresponding to a 0.68×0.82 mm^2 tissue region digitized at $5\times$ magnification.

Feature set (No. of features)	Feature description
Cytoplasm features (6)	Mean and standard deviation of L, a, b color bands in the neighborhood of a lumen point (Ω_{Lp})
Nuclei features (8)	Mean (μ) and standard deviation (σ) of L, a, b color bands in the neighborhood of a nuclei point (Ω_{Np}). μ and σ of the percentage of Ω_{Np} that contains nuclei pixels.
Lumen shape features (3)	Area, solidity (the ratio of the lumen area to its convex hull area) and circularity ((4πarea) / perimeter^2).
Morphological features (2)	Mean and standard deviation of the distance between a lumen point and a nuclei point
Contextual features (3)	Neighborhood crowdedness, shape similarity, and size similarity

Table 2.2: Features extracted for each gland segment.

classifier (RBF kernel) with this feature vector to classify the gland segment into an artifact, a normal gland or a cancer gland[3]. In the current work, all distances are measured in pixels. However, pixel distances can be converted to physical distances if the image magnification

---

[3]In the experiments (section 2.5.2), we compare the classification results of four different classifiers (SVM, Adaboost, k-nearest neighbor (k = 9), and neural network). The results obtained by SVM are better than those obtained by the other three classifiers.

is known.

2.5 Performance Evaluation

2.5.1 Gland Database

As mentioned in chapter 1, the database used in this experiment includes 48 tissue image regions at $5\times$ magnification[4] (average size is $900 \times 1,500$ pixels). Given the pathologist's annotation for each tissue image region, we manually labeled 525 artifacts, 931 normal glands and 1,375 cancer glands to form the labeled (ground truth) gland database. We also implemented the state-of-the-art methods in [2] and [3] to compare their performance with the proposed segmentation (NLA) and classification methods. The segmentation process for all the three methods starts with the lumen objects identified in each image by the k-means clustering procedure discussed in section 2.3. Since the same set of lumen objects and the same ground truth information (which is not affected by lumen objects) are used for all the three methods, the comparison among the three methods is fair.

2.5.2 Gland Classification Evaluation

We perform a 10-fold cross validation on the gland database, and report the average classification accuracy and the associated standard deviation. Since we need to extract contextual information for each gland by using the neighboring glands in the same image, we perform an image-based cross validation, i.e., we impose a constraint that glands from the same image are not included in both the training and test sets at the same time. Since artifacts can be considered noisy regions, we first solve two 2-class classification problems, i.e., (i) artifacts vs glands and (ii) normal glands vs cancer glands. Next, we perform the 3-class classification by combining the previous two 2-class classification problems in a hierarchical fashion.

---

[4]Images in some of the figures in this chapter are shown at $20\times$ for visualization purposes.

2.5.2.1 Classifier Selection

We compare the classification results of four popular classifiers: SVM [2, 41, 56], k-nearest neighbor [56, 59], Adaboost [45], and feedforward neural network. For each classifier, we evaluate different parameter values and report the best results in Table 2.3. From this table, we can see that SVM outperforms the other three classifiers. Henceforth, we use SVM as the classifier in our method.

Classification problem	Adaboost (with decision stump as weak classifier)	K-nearest neighbor (k=9)	Feedforward neural network (1 hidden layer)	SVM (RBF kernel)
Artifact vs gland	0.91 (0.03)	0.91 (0.05)	0.93 (0.03)	**0.94** (0.04)
Normal vs cancer	0.75 (0.08)	0.76 (0.08)	0.76 (0.07)	**0.79** (0.08)
All three classes	0.72 (0.06)	0.73 (0.07)	0.75 (0.06)	**0.77** (0.07)

Table 2.3: Gland classification accuracies and standard deviations (using 10-fold cross validation) obtained by different classifiers. The bold entry in each row indicates the highest accuracy for the corresponding classification problem.

2.5.2.2 Comparisons with Published Studies

In Table 2.4, besides the methods in [2], [3] and the proposed method (denoted by SVM-SCF, i.e. applying SVM classifier on the structural-contextual features (SCF)), we also report the results of the SVM-SF method (applying SVM classifier on the structural features (SF)), the results of the PPMM-SF method (applying the PPMM (section 2.2.2) on the SF), and the results of the PPMM-SCF method (applying the PPMM on the SCF). Since artifacts were addressed in [2] but not addressed in [3], we only report normal vs cancer result for the method in [3]. From Table 2.4, we can see that:

1. The proposed SVM-SCF method obtains the highest accuracy in all the classification problems.
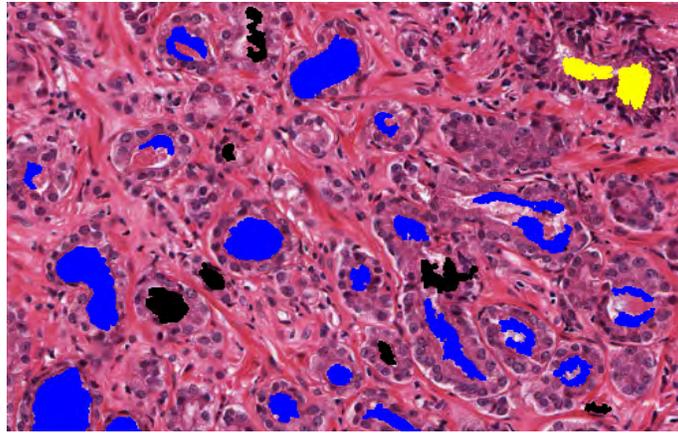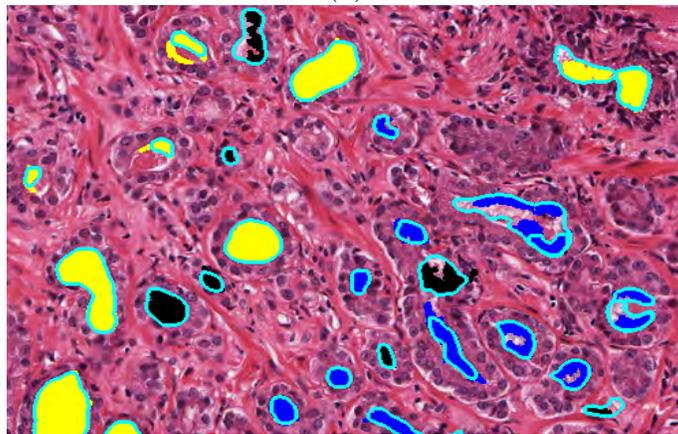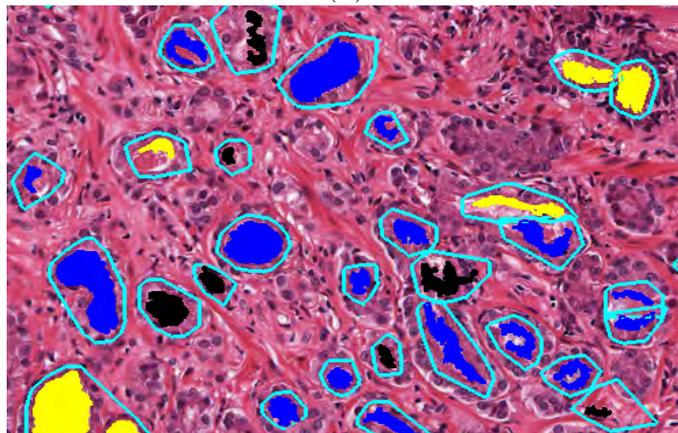
Figure 2.11: Comparison between the proposed SVM-SCF method and Monaco et al.'s method [3] for the two-class gland classification. (a) Ground truth. (b) Output of [3]. (c) Output of SVM-SCF. Cyan contours denote segmentation results, and color of the lumen corresponds to gland label (black, red, yellow and blue denote non-labeled glands, artifacts, normal glands and cancer glands, respectively). The image size is 300×500 pixels, corresponding to a 0.6×1 mm^2 tissue region digitized at $5\times$ magnification.
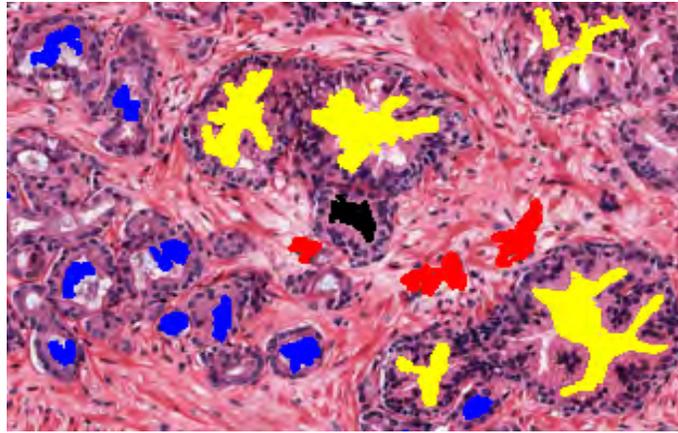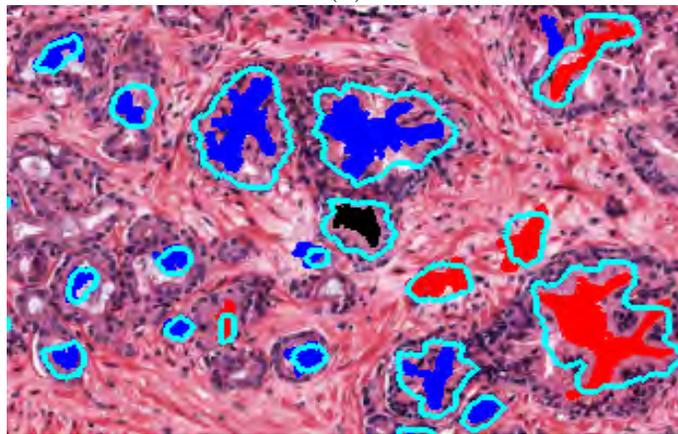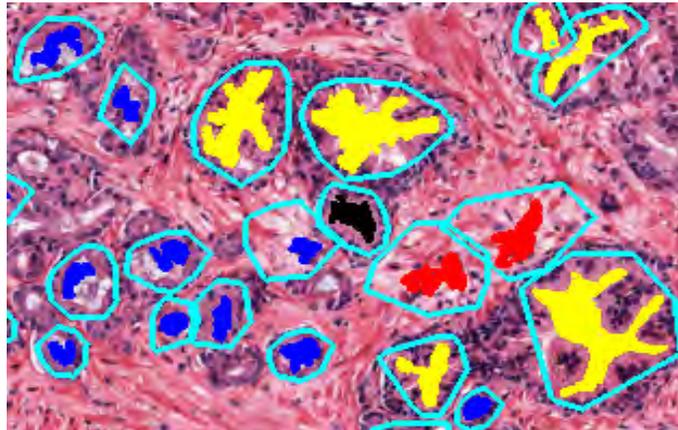
Figure 2.12: Comparison between the proposed SVM-SCF method and Naik et al.'s method [2] for the three-class gland classification. (a) Ground truth. (b) Output of [2]. (c) Output of SVM-SCF. Cyan contours denote segmentation results, and color of the lumen corresponds to gland label (black, red, yellow and blue denote non-labeled glands, artifacts, normal glands and cancer glands, respectively). The image size is 250 × 300 pixels, corresponding to a 0.5 × 0.6 mm² tissue region digitized at 5× magnification.

43

2. The superior performance of PPMM-SF over [3] shows that the structural features (SF) are useful.

3. A drawback of the PPMM is that it requires a density estimation $p(y|x)$ (y is the feature vector, and x is the class label), which is difficult when y is a high dimensional vector like the 19-dimensional SF vector or the 22-dimensional SCF vector. Here, we address this "curse of dimensionality" problem (which was not discussed in [3]) by testing three different methods to estimate $p(y|x)$, i.e., (i) Parzen window density estimator (PWDS), Naive Bayes model under the assumption that each feature is independent and follows a (ii) Gaussian distribution, and (iii) Gamma distribution (as used in [3]). The best results among these methods are obtained by PWDS, which are reported in Table 2.4 (columns 5 and 6).

4. By including artifacts in the classification procedure, we can identify them more accurately than [2], in which artifacts were identified in the pre-processing step[5], since a rich set of features are used.

5. The normal vs cancer gland classification, as expected, is more challenging than the artifact vs gland classification, i.e., there are several normal glands whose features are similar to cancer glands and vise versa. Figure 2.13 shows examples of these situations. In Figure 2.13a, a normal gland is misclassified as a cancer gland because there are weak nuclei (one layer of nuclei) on the gland boundary. In Figure 2.13b, a normal gland is misclassified as a cancer gland because the detected lumen is small. On the other hand, the cancer gland in Figure 2.13c is misclassified as normal because the gland has a large lumen.

Examples of classification outputs for the three methods (proposed method, method in [2], and method in [3]) are shown in Figures 2.11 and 2.12. In these figures, we compare the

---

[5]In [2], artifacts were discarded from true lumina by computing the lumen likelihood, cytoplasm likelihood and size likelihood of the candidate lumina. Gland segmentation was then initiated from true lumina.
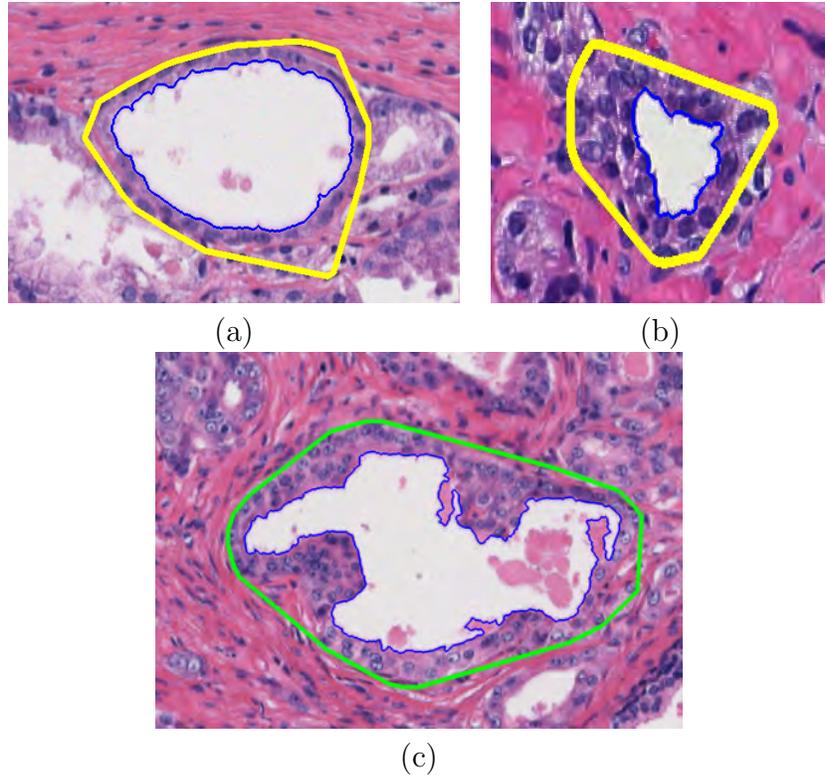
Figure 2.13: Examples of gland misclassification by the proposed method. (a) and (b) Normal glands are misclassified as cancer glands. (c) A cancer gland is misclassified as a normal gland. Yellow curve denotes normal glands and green curve denotes cancer glands.

Classification problem	Method in [2]	Method in [3]	SVM-SF	PPMM-SF	PPMM-SCF	SVM-SCF
Artifact vs true gland	0.78 (0.09)	-	0.93 (0.03)	0.93 (0.06)	0.93 (0.04)	**0.94** (0.04)
Normal vs cancer	0.67 (0.13)	0.68 (0.13)	0.75 (0.07)	0.73 (0.11)	0.75 (0.11)	**0.79** (0.08)
All three classes	0.54 (0.12)	-	0.74 (0.06)	0.75 (0.10)	0.73 (0.08)	**0.77** (0.07)

Table 2.4: Gland classification accuracies and standard deviations for methods in [2], [3], SVM-SF, PPMM-SF, PPMM-SCF, and SVM-SCF by 10-fold cross validation. The bold entry in each row indicates the highest accuracy for the corresponding classification problem.

performance of the proposed method for three-class classification result with [2] and the normal vs cancer classification result with [3].
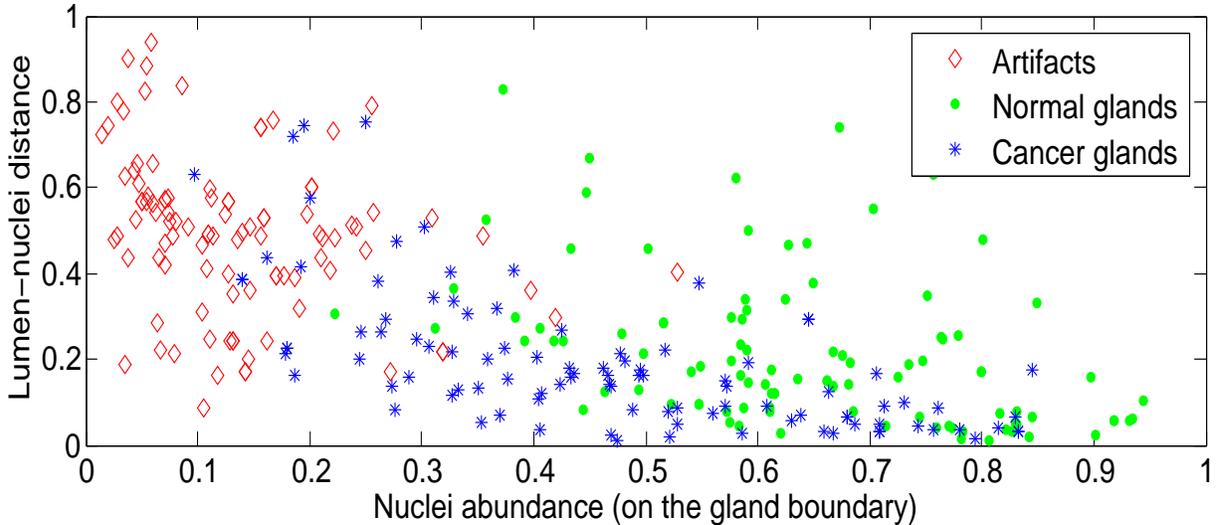
Figure 2.14: Plot of the normal glands, cancer glands and artifacts in the space of the two best features. 100 randomly selected samples from each class are plotted here.

2.5.2.3 Relaxation Labeling

Relaxation labeling [60, 61] is a well-known method in computer vision to utilize the contextual information. To compare the proposed SVM-SCF method with relaxation labeling, we implement the procedure described in [60], where objects are individual glands and the local measurements (object features) comprise the structural features. In this relaxation labeling method, which is an iterative process, every gland is assigned a confidence score, indicating how likely it is that the gland is a cancer gland. The confidence score is computed based on the local measurements of the glands, and also based on the confidence score of its neighbors [6]. However, we obtain lower accuracies for the relaxation labeling method compared to the proposed SVM-SCF method. The reason is that, when classifying a gland g, the relaxation labeling method does not take into account the number of glands in the neighborhood of g, which is an important feature to discriminate normal vs cancer glands, as the SVM-SCF method.

---

[6]To find neighbors of the glands in this method, we use the same distance threshold that is used to group glands when computing the proposed contextual features.

2.5.2.4   Feature Weights

We apply linear SVM to compute feature weights, and find the two best features: (i) nuclei abundance (on the gland boundary) and (ii) lumen-nuclei distance (Table 2.2). A plot of samples of the three classes in this 2-dimentional feature space (Figure 2.14) shows that the nuclei abundance of glands is higher than artifacts, and the nuclei abundance of normal glands is higher than cancer glands, which demonstrates the distinctiveness of the nuclei features.

Finally, we test the proposed method on the same 48 images but at $20\times$ magnification. Similar classification accuracies as $5\times$ magnification are obtained, showing that the proposed method does not require images at a very high resolution.

2.5.3   Model-free vs Model-based Gland Segmentation

By assuming that lumen is the central component of the gland, the goal of the gland segmentation procedure is to find nuclei associated with the lumen (nuclei belonging to the glands). This is feasible because lumen and nuclei can be detected from the k-means clustering algorithm mentioned in section 2.3. The gland segmentation is an important step that leads to the computation of the structural features of the glands (Table 2.2), which are then used for gland classification.

We analyze two types of gland segmentation methods in performing this task, the model-free method and the model-based method. The proposed NLA method, a model-free method, and the level set [62], a model-based method, are taken as examples (Figure 2.15). We give a brief review of the two methods as follows.

For the NLA method, we search for the nuclei surrounding the lumen. There are two important parameters in this method, namely the number of lumen-points sampled on the lumen boundary, n_{lp}, and the angle α_0 of the search region (see Figure 2.7). These two parameters determine the coverage and the fineness of the search space as well as the computation cost of the method. For all the experiments in the thesis, we empirically select
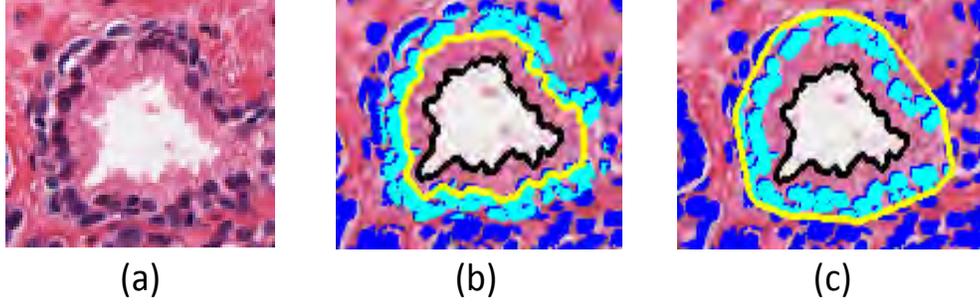
Figure 2.15: Model-based and model-free methods for gland segmentation. (a) Input gland (image size is 150 × 140 pixels, corresponding to a 0.07 × 0.07 mm^2 tissue region digitized at 20× magnification). (b) Segmentation obtained by the level set method (a model-based method). (c) Segmentation obtained by the NLA method (a model-free method). The blue regions denote all nuclei detected in the image, while the cyan regions denote the nuclei belonging to the gland found by the segmentation methods. The black contour denotes the lumen boundary, where the level set curve is initialized. The yellow contour in (b) denotes the level set curve after the evolution, while the yellow contour in (c) denotes the segmented region obtained by the NLA method.

$n_{lp} = n/3$ (n is the total number of points on the lumen boundary) and the angle $\alpha_0 = \pi/6$, which makes the search space relatively fine and cover the complete area around the lumen. The search process at every lumen-point stops when nuclei are found for the lumen-point or when a maximum search radius is met.

For the level set method, a level set curve is initialized at the lumen boundary and iteratively evolved by minimizing an energy functional until the predefined stopping criteria are met. As described in [62], the energy formulation of the level set includes three terms:

$$\mathcal{E}(\phi) = \mu\mathcal{R}(\phi) + \lambda\mathcal{L}_g(\phi) + \alpha\mathcal{A}_g(\phi) \tag{2.9}$$

where $\mathcal{R}(\phi)$ denotes the regularization term, $\mathcal{L}(\phi)$ denotes the energy functional term that drives the curve to the nuclei regions (the external energy), and $\mathcal{A}(\phi)$ denotes the energy functional term that speeds up the evolution process. As mentioned in [62] and by empirical experiments, the two most important parameters are α (the coefficient of the term $\mathcal{A}_g(\phi)$) and C_0 (the initial value of the level set function). To choose the stopping criteria, we follow the suggestion in [62]: the curve is evolved until (i) it does not change for T_1 consecutive iterations or (ii) the number of iterations exceeds a threshold T_2. We use $T_1 = 5$ (as suggested

48

Proposed NLA method	Level set method [62]
1. The two most important parameters are: the number of lumen-points and the angle of the search region (see Figure 2.7)	1. The two most important parameters are: α (see Equation 2.9) and C_0 (the initial value of the level set function).
2. Stops when nuclei are found; uses the MAD measure to discard outliers	2. Stop after T_2 iterations or stop when the curve does not change for T_1 consecutive iterations
	3. The initial level set curve can be split into multiple curves during the evolution (Figure 2.16)

Table 2.5: Comparison of the proposed NLA and the level set methods for gland segmentation.

in [62]) and $T_2 = 600$, which is a very large number (the values of T_2 used in [62] are less than 300). By using a large value of T_2, we expect the curve to capture the glands with complex shapes despite its large computation time. Table 2.5 summarizes the differences between the two methods.

To obtain nuclei belonging to the gland in the level set method, we first sample a set of points on the level set curve (after the evolution) and consider these points as nuclei points (which are similar to the nuclei points mentioned in the NLA method (Algorithm 2.1)). The nuclei belonging to the glands are the nuclei falling within the neighborhood of these nuclei points (the cyan blobs in Figure 2.15b). The size of the neighborhood is chosen to be the same as the NLA method.
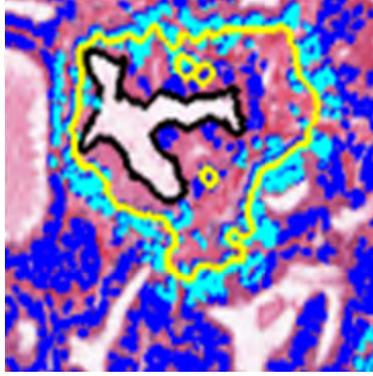
Figure 2.16: The level set curve (yellow) is split into multiple curves during the evolution, which is an unexpected behaviour.

We now demonstrate segmentation examples of the two methods. Figure 2.17a is an example of a cancer region in which glands have very few nuclei on the boundary. In this case, the proposed NLA method does not provide a good segmentation result because the search region covers the distant nuclei (Figure 2.17b). On the other hand, in Figures 2.17d and 2.18a where glands contain a large number of nuclei on the boundary, the NLA method provides good segmentation results because the search region only captures these nuclei (Figures 2.17e and 2.18b).

For the level set method, depending on the choice of parameters, there are two types of behaviour of the curve evolution:

1. The curve evolution is conservative, i.e., low freedom of curve evolution. We refer to the parameter values corresponding to this type of behaviour as *parameter choice 1*.

2. The curve evolution is progressive, i.e., high freedom of curve evolution. We refer to the parameter values corresponding to this type of behaviour as *parameter choice 2*.

When *parameter choice 1* is used, the level set curve tends to maintain a compact shape, making it suitable to segment small glands with compact, circular shape, and containing only a few nuclei on the boundary (these glands mostly appear in cancer tissue regions). This can be illustrated in Figure 2.17c. However, this parameter choice prevents the curve from capturing the glands with complex, irregular shapes and with a large variation in size
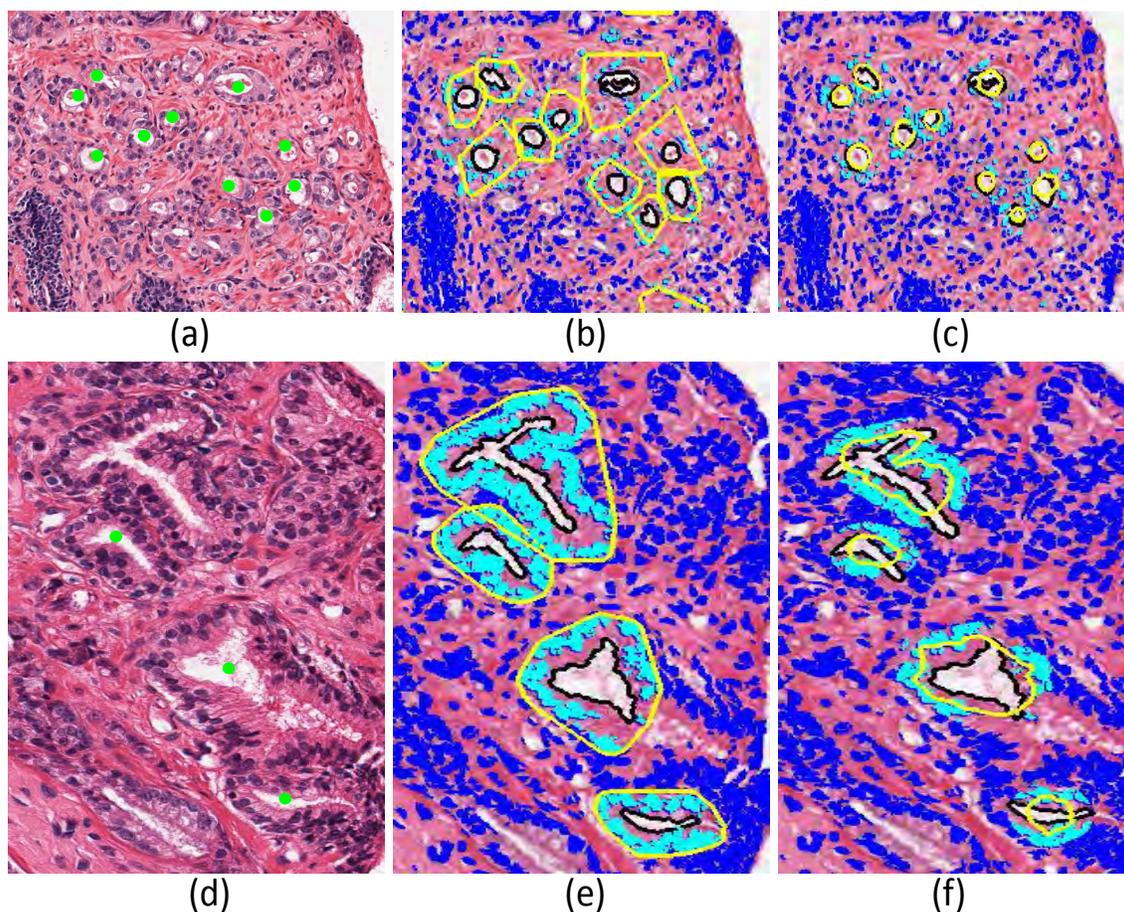
Figure 2.17: Examples of gland segmentation results, in which the level set curve has low freedom of evolution (parameter choice 1). (a) A cancer tissue image in which glands have only a few nuclei on the boundary (image size is 230×250 pixels, corresponding to a 0.46×0.5 mm^2 tissue region digitized at $5\times$ magnification). (b) and (c) Segmentation results of the NLA method and the level set method for the image in (a). (d) A normal tissue image in which glands have abundant nuclei on the boundary (image size is 180×150 pixels, corresponding to a 0.36×0.3 mm^2 tissue region digitized at $5\times$ magnification). (e) and (f) Segmentation results of the NLA method and the level set method for the image in (d).
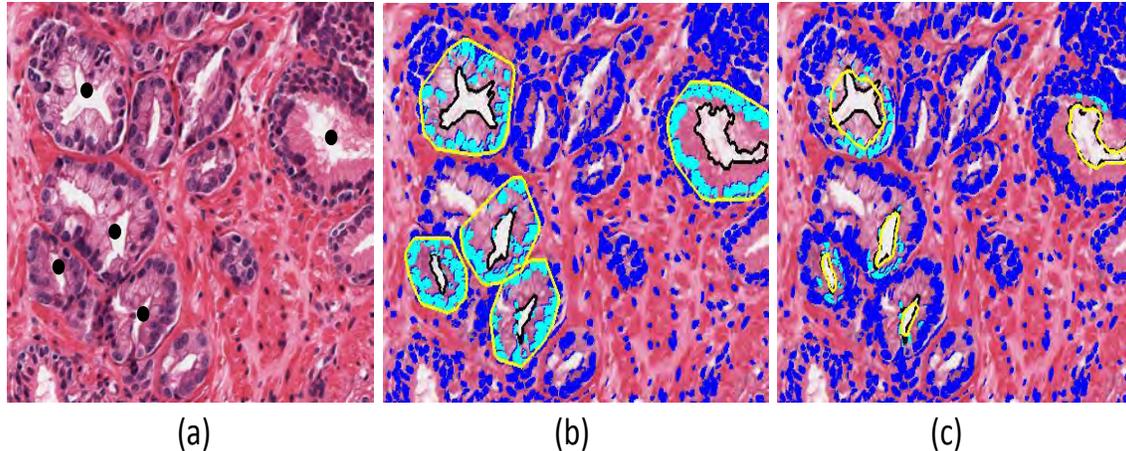
Figure 2.18: Examples of gland segmentation results, in which the level set curve has low freedom of evolution (parameter choice 1). (a) A cancer tissue image in which glands have abundant nuclei on the boundary (image size is 150×170 pixels, corresponding to a 0.3×0.34 mm^2 tissue region digitized at $5\times$ magnification). (b) and (c) Segmentation results of the NLA method and level set method for the image in (a).

(Figure 2.17f)[7], which are very typical in normal tissue regions. Moreover, the parameter choice 1 also makes the curve evolve very slowly (Figure 2.18c shows the result of the curve evolution after 300 iterations).

On the other hand, when *parameter choice 2* is used, the level set curve is able to capture the complex shape of the glands (Figure 2.19c). However, this freedom in evolution may make the curve fail to converge to the nuclei surrounding the glands when the glands have only a few nuclei on the boundary (Figure 2.19f)[8].

Since the final goal is to classify glands, we are interested in comparing the gland classification results derived by the segmentation results of the level set method and of the NLA method (the gland classification results derived by the NLA method were reported in the last column of Table 2.4). To evaluate the gland classification results derived by the level set method, we compute the same 22 features described in Table 2.2 for the gland segments obtained by the level set method, and use the same experimental setup described in

---

[7]We also tried a larger threshold value $T_2 = 900$ iterations, however, the level set curve still cannot capture the complete gland region as the NLA method.

[8]We also tried a smaller threshold value $T_2 = 300$ iterations, however, the curve still does not converge to the nuclei surrounding the gland.
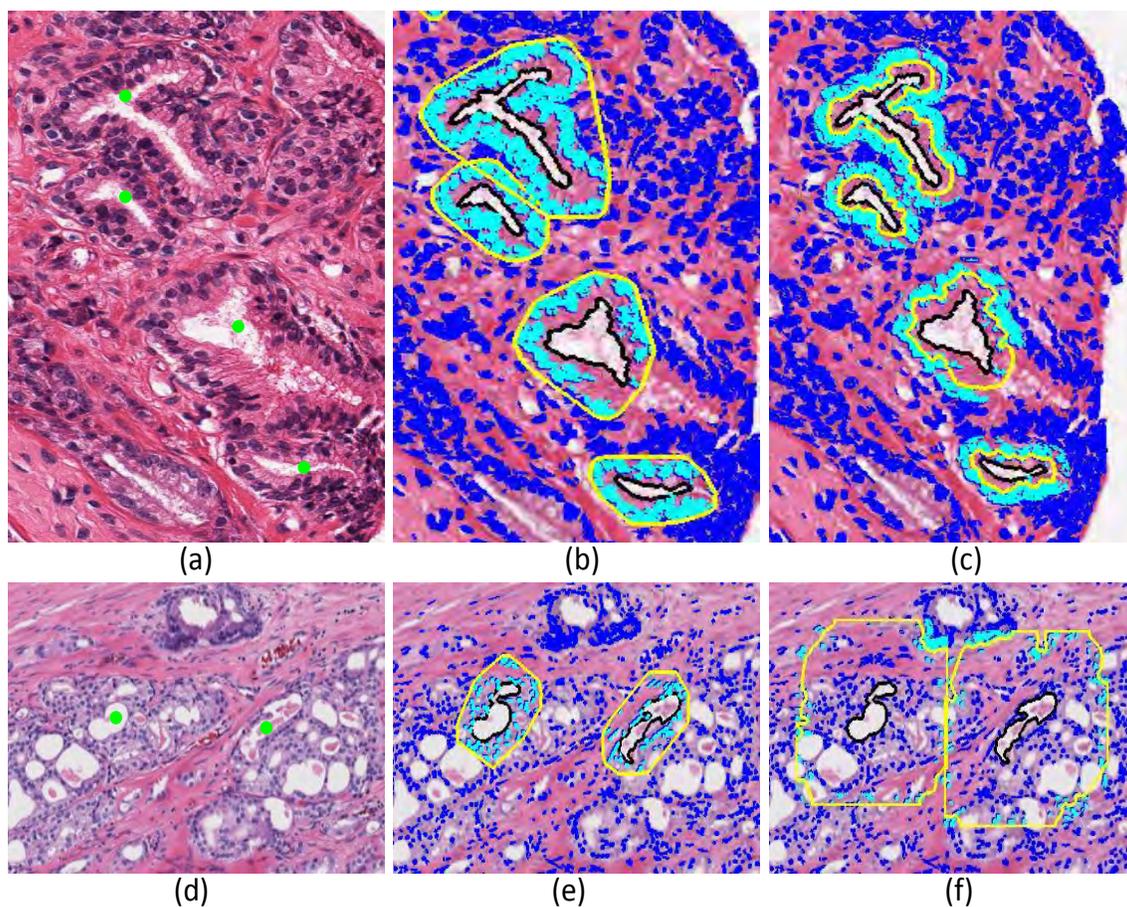
Figure 2.19: Examples of the gland segmentation results, in which the level set curve has high freedom of evolution (parameter choice 2). (a) A normal tissue image in which glands have abundant nuclei on the boundary (image size is 180×150 pixels, corresponding to a 0.36×0.3 mm^2 tissue region digitized at $5\times$ magnification). (b) and (c) Segmentation results of the NLA method and level set method for the image in (a). (d) A cancer tissue image in which glands have only a few nuclei on the boundary (image size is 200×220 pixels, corresponding to a 0.4×0.44 mm^2 tissue region digitized at $5\times$ magnification). (e) and (f) Segmentation results of the NLA method and level set method for the image in (d).

Classification problem	Level set method	NLA method
Artifact vs gland	0.93 (0.09)	**0.94** (0.09)
Normal gland vs cancer gland	0.75 (0.06)	**0.79** (0.08)
All three classes	0.74 (0.07)	**0.77** (0.07)

Table 2.6: Gland classification accuracies (standard deviations) derived by the gland segmentation results of the level set and of the proposed NLA methods. The bold entry in each row indicates the highest accuracy for the corresponding classification problem.

section 2.5.2. By evaluating different parameter values of the level set method, we obtain the best gland classification results derived by this method in Table 2.6 (the gland classification results derived by the NLA method are also listed here). This table shows that the NLA method leads to better gland classification results than the level set method (with different parameter values evaluated).

2.5.4 Manual vs Automatic Gland Segmentation

We further evaluate the proposed NLA method by comparing it with the manual gland segmentation method in this section. Since our final goal is to classify glands, we compare the gland classification results derived by the gland segmentation results of the NLA method and of the manual segmentation method, instead of directly comparing the segmentation results. To achieve this goal, we manually segment 247 glands (including 90 normal and 157 cancer glands) from 22 different tissue images[9]. These images and glands are selected from the database of 48 images used in this chapter (section 2.5.1) in such a way that they cover most of the variations in color, size, shape, and structure of glands in this database.

In the manual gland segmentation method, we manually find the nuclei associated with the lumen objects that represent the glands. Figure 2.20 shows examples of the manual gland segmentation results, and the corresponding automatic gland segmentation results obtained by the NLA method. Given the manual gland segments, we extract the 22 features described in Table 2.2 from these segments. Next, we classify the segments into normal and cancer glands, using SVM classifier (RBF kernel) and the 10-fold cross validation technique. The same procedure is applied for the NLA method on the same 247 glands. The classification accuracies obtained by the two methods are almost the same, 0.80 (standard deviation is 0.08), illustrating that the gland segmentation result of the NLA method is comparable to the manual segmentation for classification purposes.

2.6 Summary and Contributions

In this chapter, we have addressed the gland segmentation and classification problem. The first contribution here is the gland segmentation method, termed nuclei-lumen-association (NLA), that leads to better gland classification results than the level set method and comparable gland classification results to the manual gland segmentation method.

---

[9]The manual segmentation is conducted by the author of this thesis.
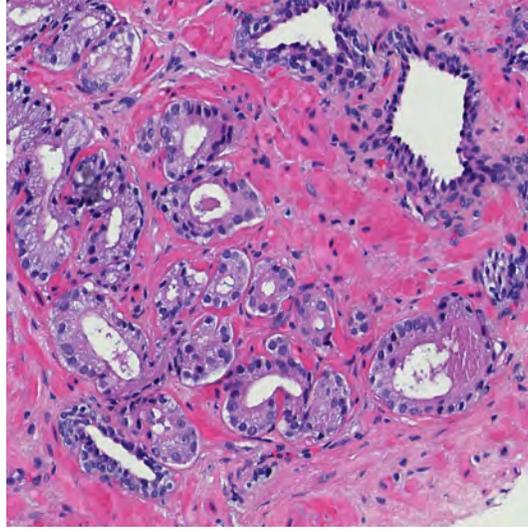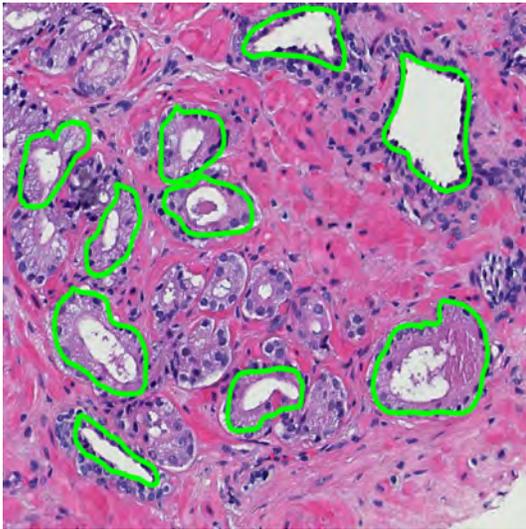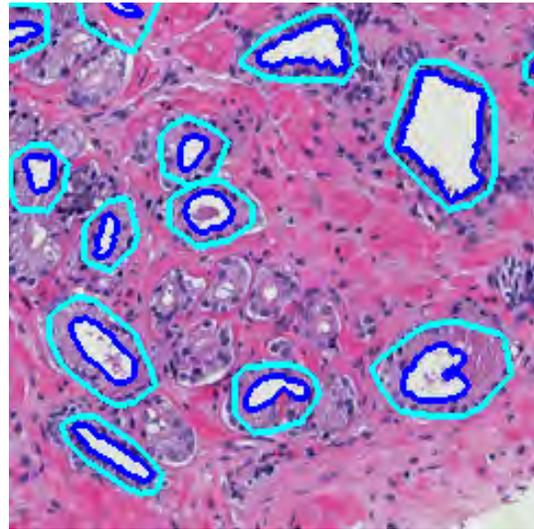
Figure 2.20: Manual vs automatic gland segmentation. (a) A tissue image of 240×235 pixels, corresponding to a 0.48×0.47 mm^2 tissue region digitized at $5\times$ magnification. (b) Manual segmentation results of the glands in (a) (green curves). (c) Automatic segmentation results by the NLA method of the glands in (a) (cyan curves).

The second contribution is a salient representation (feature vector) for the gland in the prostate tissue image. The proposed feature vector captures all the relevant information about the gland, including both structural information (information about lumen, cytoplasm and nuclei), as well as contextual information, i.e., the relation of the gland with other glands in the neighborhood area. Comparisons with several published methods were conducted to show the robustness of the extracted features.

Finally, we evaluate the gland classification performance by using a gland dataset containing three classes: artifacts, normal and cancer glands. Such a study has not been reported in the literature. This procedure allows us to precisely analyze the discriminative power of various gland features. Moreover, by including artifacts in the classification procedure instead of the preprocessing step, we identify artifacts with a higher accuracy than the published method in the literature.

Chapter 3

TISSUE IMAGE CLASSIFICATION

3.1 Introduction

In Gleason grading, the pathologist needs to assign grades to different regions in the tissue slide image. Due to the large size of the tissue slide image (as discussed in chapter 1) and the importance of the discrimination between tissue structures of different Gleason grades, an automatic tool to classify a tissue region into one of the Gleason grades would be very useful. This explains why the tissue image classification problem has received the most attention in the literature (chapter 1).

In this chapter we address the tissue image classification problem by classifying a tissue image region into one of the three classes: normal, Gleason grade 3 and Gleason grade 4, which are the three most common cases in the Gleason grading of prostate cancer. Examples of images from these three classes are shown in Figure 3.1. A description of gland structures in different Gleason grades was given in chapter 1. Here, we briefly review the properties of these classes. Glands in normal images typically have large size, large lumen, complex shape, and abundant nuclei on the boundary. In a Gleason grade 3 image, glands are generally smaller with regular shape and size, and contain less abundant nuclei on the boundary. Finally, in grade 4 images, glands usually fuse together instead of staying separated as in grade 3 images. In chapter 2, we introduced the methods to segment glands from the tissue images and extract features of the glands. We utilize these methods to solve the tissue image classification problem in this chapter.
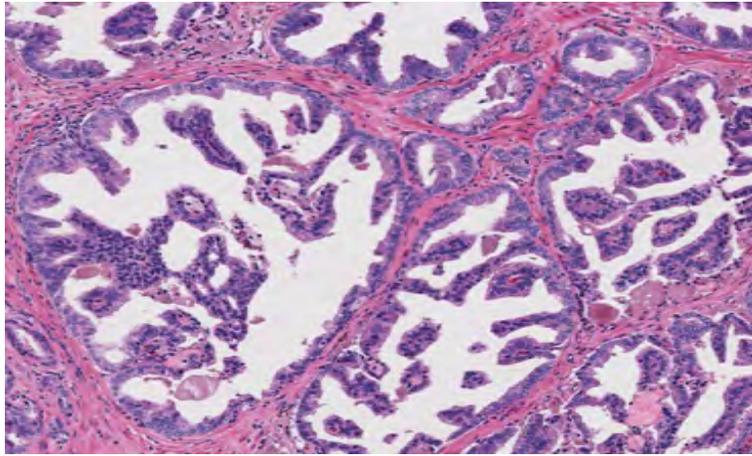
(a)
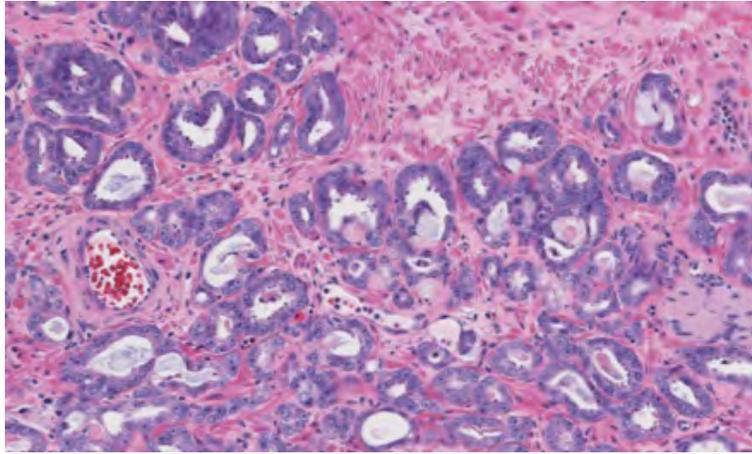
(b)
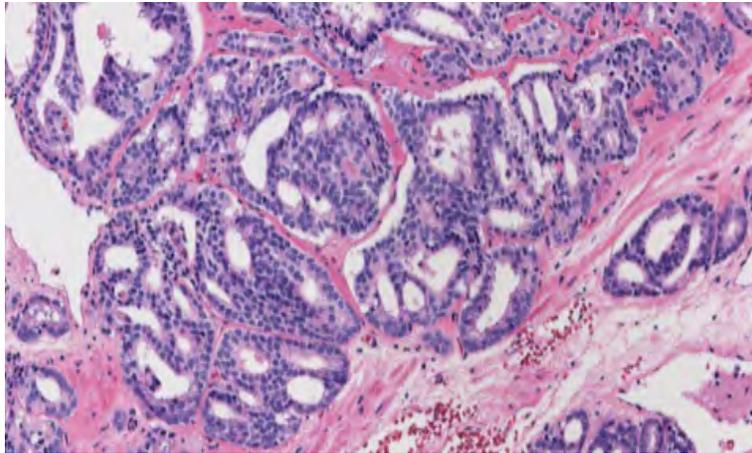
(c)

Figure 3.1: Three classes of interest: (a) Normal image, (b) Gleason grade 3 image and (c) Gleason grade 4 image. The average image size is 300×370 pixels, corresponding to a 0.6 $\times$ 0.74 mm^2 tissue region digitized at 5$\times$ magnification.

3.2 Related Work

Most studies reported in the literature have used textural features to perform the tissue image classification. Diamond et al. [63] used co-occurrence features [4] to classify each 100×100 sub-region in a tissue image into either stroma or prostatic carcinoma. In addition, lumen area was used to discriminate benign tissues from the other two classes. An accuracy of 79.3% was reported on sub-regions of 8 tissue images ($40\times$ magnification). In [41], fractal dimension features were calculated for the tissue image to discriminate the textural discrepancy between low grade and high grade carcinoma. By using an SVM classifier with leave-one-out technique, the method achieved 86.3% accuracy for the classification of 1,000 prostatic biopsy images into normal, Gleason grade 3, 4 and 5 classes. A multiwavelet transform was used as the main texture analysis tool in [59]. The features used for classification included entropy and energy derived from the multiwavelet coefficients of the image. Ten different types of multiwavelets were evaluated on a dataset of 100 prostate sample images ($100\times$ magnification) of Gleason grades 2, 3, 4, and 5. By using the leave-one-out cross validation technique for this 4-class classification problem, the authors obtained the best result of 97% for "correct classification percentage". In another study, Tabesh et al. [56] employed both global features of the entire image and local features of every segmented object in the image. Global features included color histogram, fractal features, texture and morphometry of the image. Local features were computed for histological objects such as nuclei, stroma and lumen, which were extracted by the MAGIC system [64]. They achieved 96.7% accuracy for tumor vs nontumor classification (five-fold cross validation with 367 images) and 81% accuracy for low grade vs high grade classification (five-fold cross validation with 268 images). All images used in their experiments were at $20\times$ magnification. In [40], Khurd et al. used a bag of words model for the classification of tissue images into Gleason grade 3 and Gleason grade 4 classes. They applied a bank of invariant filters [65] on the image to obtain the filter responses at each pixel. A clustering procedure was applied on the filter responses to obtain $C = 16$ clusters (textons). For each image, each pixel was assigned one

of the C textons and a histogram was computed, indicating the number of pixels in each texton. The algorithm was tested on a dataset of 25 Gleason grade 3 images and 50 Gleason grade 4 images (10× magnification), to obtain a 94% classification accuracy. A summary of these studies is reported in Table 3.1. Note that the accuracies reported for these methods cannot be directly compared because (i) different studies have used different databases, and (ii) results are for different classification problems. Further, the databases are proprietary and so they cannot be shared.

The reason that different studies used different image magnifications is primarily because (i) there is no public domain database available (as mentioned above), and (ii) the tissue slides used in different studies were digitized by different tissue slide scanners. As mentioned in chapter 1, while most tissue slide scanners can create images at 40× magnification [23,25–27], only a few slide scanners can generate 100× magnification images [28,29]. The relationship between image magnification and the classification accuracy will be analyzed later.

3.3   Proposed Method

Since images belonging to the three classes of interest differ in gland structures, it is reasonable to use *glandular features*, which are described in Table 2.2 in chapter 2 (22 structural-contextual features), to classify the images. The flowchart of the proposed tissue image classification method is illustrated in Figure 3.2. In this method, we first perform gland segmentation to identify glands in the tissue images (note that artifacts may be present in those glands). The features of these glands are also extracted, and used to identify artifacts in the images (note that these artifacts are considered as noisy regions). In order to identify artifacts, we obtain an independent gland dataset, each of which is associated with a label (artifact or true gland)[1]. An SVM classifier is trained on this gland dataset, and is used to classify all segmented glands into artifacts and true glands; artifacts are then discarded. The glandular features of the image are computed by averaging the features (Table 2.2) of

---

[1]We obtain 200 artifacts and 230 glands for this dataset.
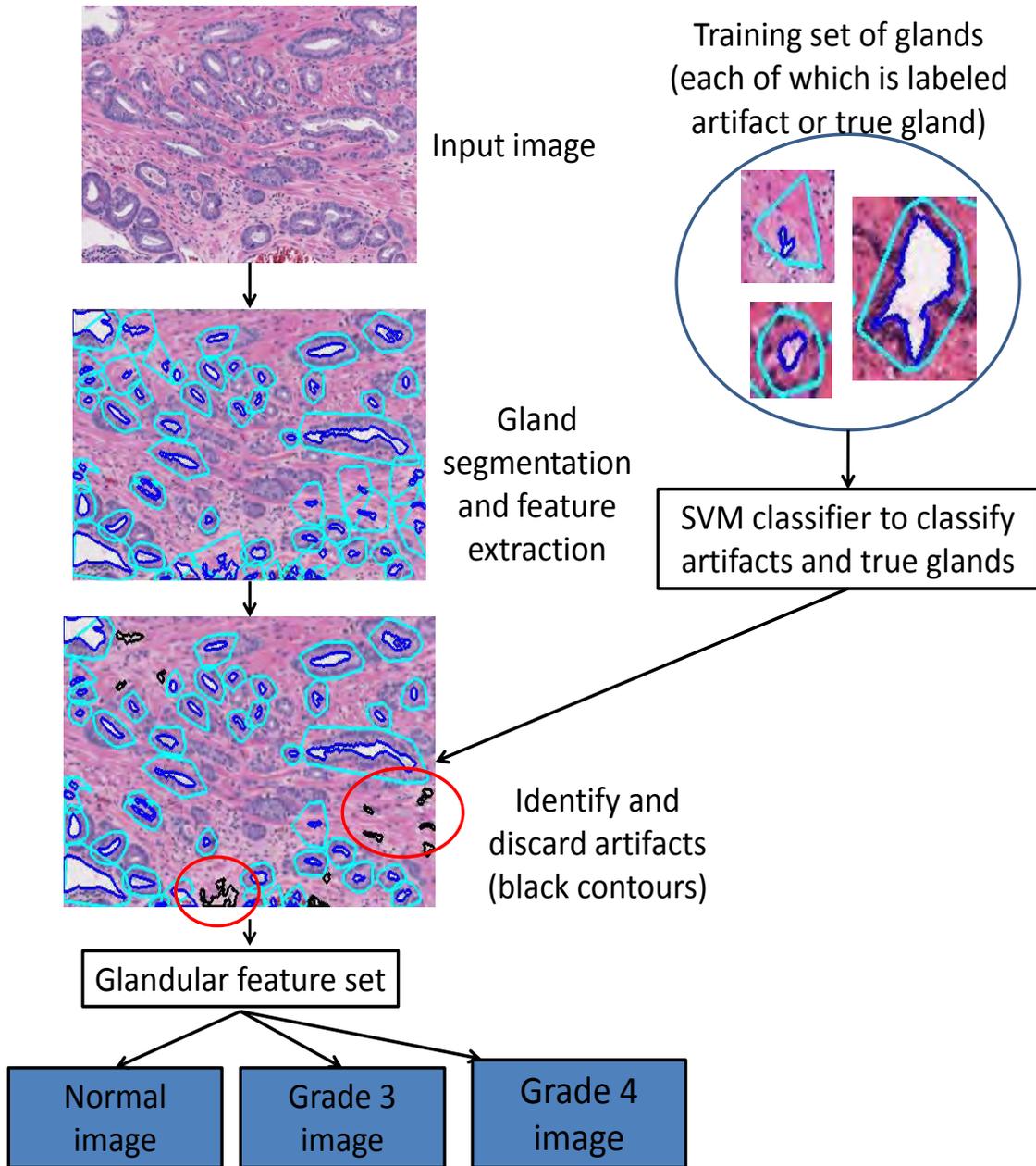
Figure 3.2: Flowchart of the proposed tissue image classification method.

Authors	Features used	Dataset size (magnification)	Classes	Accuracy
Diamond et al. [63]	Features computed from the graylevel co-occurrence matrix and lumen area	100×100 sub-regions of 8 tissue images (40×)	Stroma, benign tissue and pro-static carcinoma	79.3%
Khurd et al. [40]	Texton histogram (BoW model)	75 images (10×)	Gleason grades 3 and 4	94%
Khouzani et al. [59]	Entropy and energy of the multiwavelet coefficients	100 images (100×)	Gleason grades 2, 3, 4 and 5	97%
Tabesh et al. [56]	Global features of the image and local features of histological objects	268 images (20×) 367 images (20×)	Low grade vs High grade; Tumor vs Non-tumor	81% 96.7%
Tai et al. [41]	Fractal dimension features	1,000 images	Normal, Gleason grades 3, 4 and 5	86.3%
Proposed method	Glandular features	317 images (20×)	Normal, Gleason grades 3 and 4	83.3%

Table 3.1: Summary of prostatic tissue image classification studies related to prostate cancer. Different terminologies which have the same meaning were used in the related work. Cancer (with different Gleason grades), tumor and carcinoma refer to the tissues which are detected to have malignant properties of a cancer (cells grow aggressively, invade the surrounding tissues and spread to the non-adjacent tissues). Normal, benign and nontumor refer to the tissues which are not cancerous.

the true glands in the image, which are then used to classify the image into normal, Gleason grade 3 and Gleason grade 4. For the grade 3 vs grade 4 classification, we do not use the contextual features because glands in both grade 3 and grade 4 images are cancer glands, thus, they have similar contextual properties (in close proximity of each other and forming groups). Similar to the experiments mentioned in section 2.5.2, we also use an SVM classifier (with RBF kernel) to perform this classification.

3.4 Evaluation

3.4.1 Database

As mentioned in chapter 1, the database we use in this experiment includes 317 images (including 113 normal, 134 grade 3 and 70 grade 4)[2] at both 20× and 5× magnification (average image size at 20× is 1,400 × 1,380 pixels), each of which represents one tissue pattern corresponding to one of the three classes of interest. The grade of each image was determined by a pathologist. We perform 10-fold cross-validation and compute the average classification accuracy to evaluate the proposed method. Examples of these images are shown in Figure 3.1.

3.4.2 Classification Results

To show the robustness of the proposed method, we compare it with four different texture-based methods (i) GLCM: features computed from the gray-level co-occurrence matrix [4], (ii) BoW-filter: bag of words model and the maximum response filters [40], (iii) BoW-SIFT: bag of words model and the scale-invariant feature transform (SIFT) algorithm [66], and (iv) the spatial pyramid matching method [67][3]. The GLCM method (i) is a very popular method for texture analysis in histological images [68–72], while the BoW-filter method (ii) was proposed in [40] as an application of the popular BoW model in prostate tissue image analysis. Finally, the methods (iii) and (iv) are popular methods in computer vision.

Besides the three-class classification problem, we also report results of the following three two-class classification problems: normal vs grade 3; normal vs grade 4; and grade 3 vs grade 4. To investigate the relationship between image magnification and classification performance, we evaluate the five methods (GLCM, BoW-filter, BoW-SIFT, spatial pyramid matching, and the proposed method) using the same 317 images in the database at both

---

[2]Each image corresponds to a tissue image region in the main database mentioned in chapter 1.

[3]The spatial pyramid matching method also uses SIFT to compute the features.

Method	All three classes		Normal vs Grade 3		Normal vs Grade 4		Grade 3 vs Grade 4	
	5×	20×	5×	20×	5×	20×	5×	20×
Proposed method	**87.1** (5.7)	**87.0** (4.9)	**97.2** (3.8)	**96.7** (4.1)	**98.9** (2.3)	**98.3** (2.7)	**84.0** (6.2)	**83.3** (7.1)
GLCM	79.9 (7.6)	79.2 (8.1)	92.1 (4.7)	93.1 (6.5)	95.1 (3.1)	94.5 (3.3)	76.6 (11.0)	79.9 (9.4)
BoW-filter	80.2 (8.0)	79.8 (8.4)	96.0 (3.3)	95.4 (6.4)	97.4 (3.6)	97.0 (6.1)	75.8 (10.7)	77.3 (8.2)
BoW-SIFT	78.6 (8.7)	79.2 (5.7)	92.2 (4.4)	91.8 (5.8)	95.8 (5.3)	97.7 (3.2)	76.2 (11.1)	80.0 (9.1)
Spatial pyramid matching	77.6 (7.9)	77.6 (7.7)	93.5 (5.0)	92.9 (5.2)	94.6 (5.7)	94.2 (5.2)	75.0 (10.9)	78.1 (9.0)

Table 3.2: Average 10-fold cross validation accuracy (%) and standard deviation obtained by the proposed method, the GLCM, BoW-filter, BoW-SIFT, spatial pyramid matching methods for various tissue image classification problems on both 5× and 20× magnification images. The bold entry in each column indicates the highest accuracy for the corresponding classification problem.

5× and 20× magnification (see Table 3.2). Based on these results, we make the following observations.

- The proposed method obtains the best results for both 5× and 20× images.

- For normal vs cancer (grade 3 and grade 4) classification, for all the methods, using high magnification images (20×) does not help to improve the accuracy compared to using low magnification images (5×). Since the gland structures in normal and cancer images are very different from each other, the feature vectors extracted from low magnification images provide sufficient discrimination.

- For grade 3 vs grade 4 classification, the differences between these two grades are more subtle than the differences between normal and cancer. For all the four texture-based methods, using 20× images leads to a better accuracy than using 5× images since the texture information is richer at higher magnification images. In contrast, the proposed method obtains comparable accuracies for grade 3 vs grade 4 classification on both 5× and 20× images. The reason is that the structural features used by the proposed

method mostly rely on the information extracted from the nuclei and lumen (size and shape of the lumen, color and density of nuclei, distance between lumen and nuclei). By using the color-based pixel clustering procedure to detect lumen and nuclei, which have distinctive colors (nuclei are blue and lumina are white), their detection results are not significantly affected by the image magnification. In Figure 3.3, we do not see any significant difference between the nuclei and lumen detection results in $20\times$ and $5\times$ images.

- The performance of the spatial pyramid matching method is generally lower than that of the BoW-SIFT method (both methods use the same SIFT features).

The above empirical analysis suggests using the proposed method on $5\times$ magnification images, which requires low computational cost, for the tissue image classification problem.

Examples of correct classifications and misclassifications by the proposed method are shown in Figure 3.4 and Figure 3.5, respectively. The normal image in Figure 3.5a is misclassified as Gleason grade 3 because of the presence of (i) some small glands resembling cancer glands, and (ii) few nuclei on the gland boundary. Further, the color information is not sufficiently discriminative (cytoplasm and stroma have very similar color). The Gleason grade 3 image in Figure 3.5b is misclassified as normal image because (i) there are some normal glands present in the image (yellow rectangles), (ii) most glands have large area comparable to normal gland area and (iii) the color information is not discriminative. Finally, the Gleason grade 4 image in Figure 3.5c is misclassified as Gleason grade 3 because the cancer glands in this image are in the intermediate stage of evolving from grade 3 to grade 4, with some glands still well separated from neighboring glands.
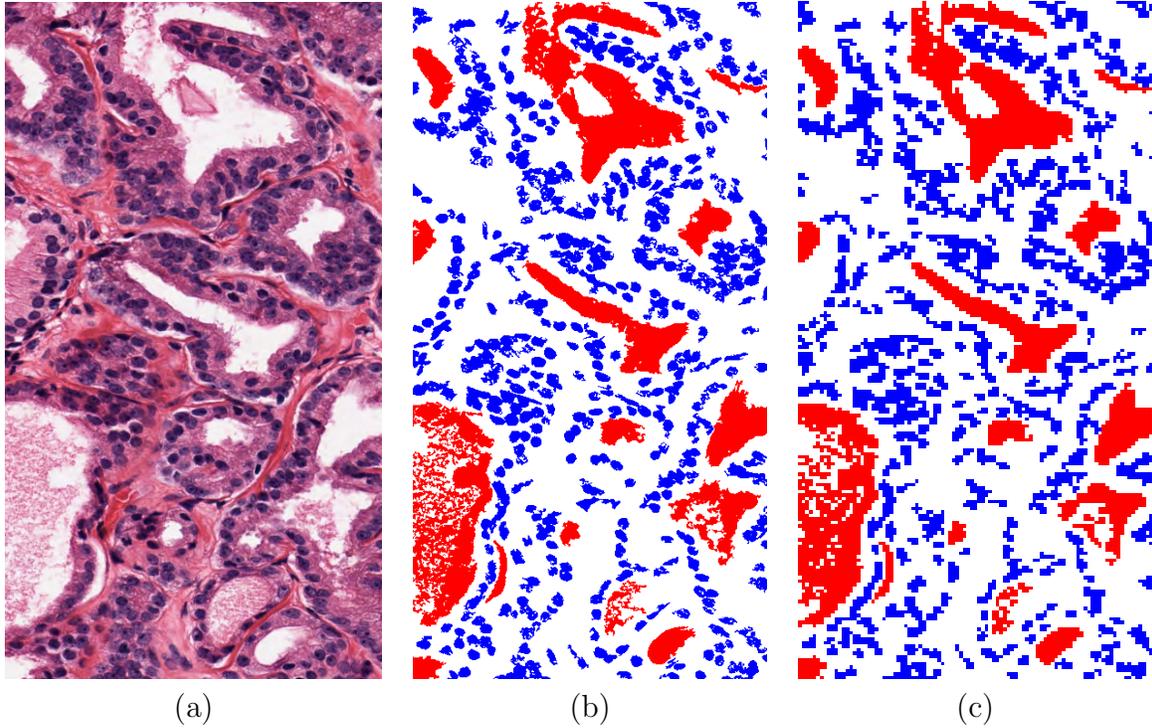
Figure 3.3: The lumen and nuclei detection results for an image at both 20× and 5× magnification. (a) A grade 3 tissue image of 790 × 390 pixels, corresponding to a 0.4 × 0.2 mm^2 tissue region digitized at 20× magnification. The lumen and nuclei detected for the (b) 20× image and (c) 5× image, where red indicates lumen and blue indicates nuclei.

3.4.3   Feature Averaging vs Per-Gland Classification

In chapter 2, we have performed the individual gland classification problem. Hence, in this section, we are interested in reusing the individual gland classification scheme in classifying the tissue image, leading to a method called *per-gland classification.* Instead of using the averages of the feature values of the glands to classify the image, the per-gland classification method performs the following steps.

1. Classifies every gland in the test image into normal, grade 3 or grade 4. For this classification, we obtain the training glands from the training images, and assign these glands the same labels as the training images they belong to.

2. Uses a majority voting scheme to determine the label of the test image, i.e., the predicted label of the test image is the dominant gland label in the image.
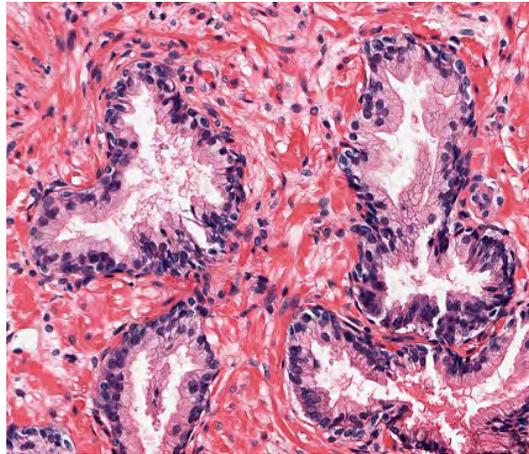
(a)
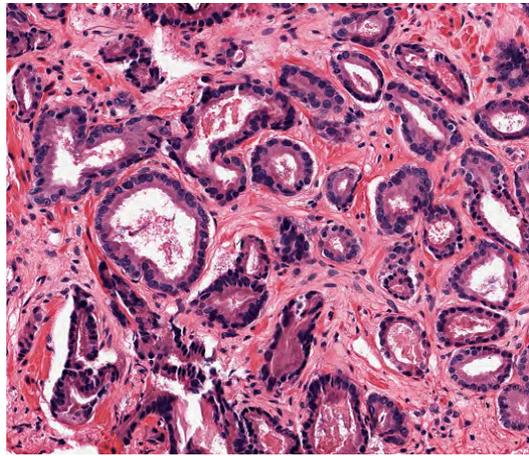
(b)
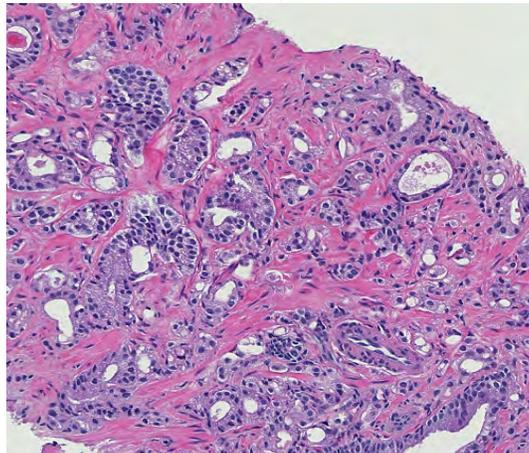
(a)

Figure 3.4: Examples of correct classification. Normal image (a), Gleason grade 3 image (b), and Gleason grade 4 image (c) that are correctly classified by the proposed method. The average image size is 250×270 pixels, corresponding to a 0.5×0.54 mm^2 tissue region digitized at $5\times$ magnification.
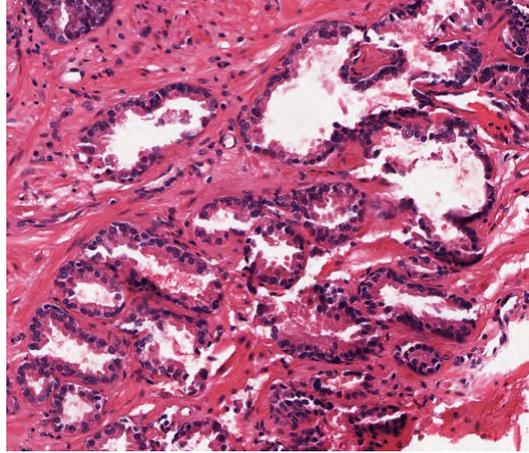
(a)
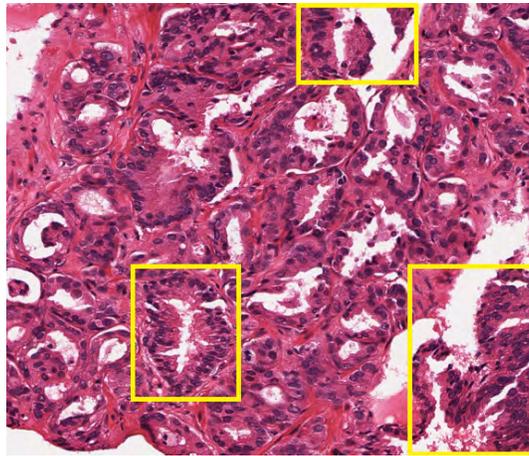
(b)
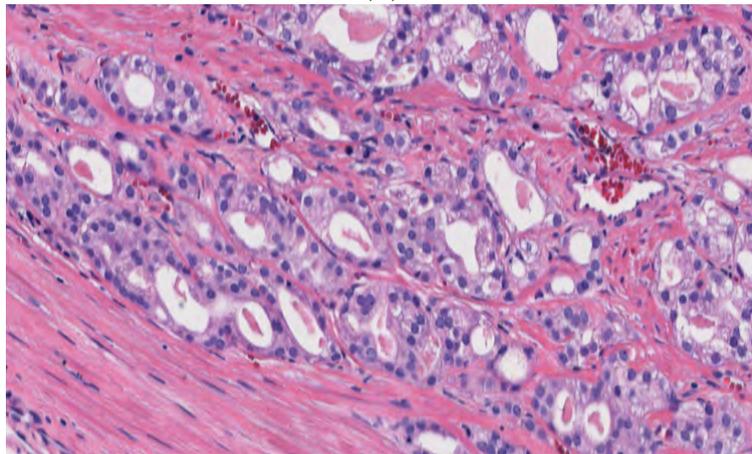
(c)

Figure 3.5: Examples of images that have been misclassified. (a) A normal image being misclassified as Gleason grade 3, (b) a Gleason grade 3 image being misclassified as normal, and (c) a Gleason grade 4 image being misclassified as Gleason grade 3. The average image size is 350×400 pixels, corresponding to a 0.7×0.8 mm^2 tissue region digitized at $5\times$ magnification.

The classification results obtained by this per-gland classification method are lower than those of the method based on averaging of gland features. The reasons for this can be explained as follows.

1. An image may contain glands that have different grades from the image (e.g., a grade 4 image may contain normal glands or grade 3 glands), which can be considered as noisy glands. Hence, we have more confidence in the grade of the entire image than the grades of individual glands. As a result, the training data that include (feature vector of the image - average of gland features in the image, image label) are more reliable than the training data that include (feature vector of the gland, gland label). Moreover, by computing the averages of the gland features in the image, the features of the glands that actually represent the image dominate the features of the noisy glands (since these noisy glands appear less common in the image).

2. For the per-gland classification method, to predict the label of the image, we have to predict the label of each individual gland in the image. Since each gland label prediction has an associated error rate, the combination of these per-gland prediction results will accumulate the error rates.

As a result, using the average of glandular features is a better choice for the tissue image classification problem.

Although the proposed method obtains very good results for normal vs cancer (grade 3 and grade 4) classification, the result for grade 3 vs grade 4 classification is not as good. In the next section, we introduce a new gland segmentation method, termed *nuclei-based gland segmentation*, and a method to compute the gland-score of a segment to indicate how similar the segment is to the gland. Next, we describe how to use these methods to improve the grade 3 vs grade 4 classification result.

Method	Description
Level set [2]	Initialize the level set curve at the lumen boundary and evolve the curve to the nuclei area
Region growing [3]	Select a seed in the lumen region and grow the gland region from the seed
Proposed NLA	Sample the lumen-points on the lumen boundary and find nuclei associated with the lumen

Table 3.3: Summary of state-of-the-art gland segmentation methods, which are considered as lumen-based methods.

3.5 Nuclei-based Gland Segmentation

Limitation of the Lumen-based Gland Segmentation Methods:

State-of-the-art gland segmentation methods (including the proposed NLA method), mentioned in chapter 2, Table 2.1, rely mainly on lumen: they consider the lumen as the center of the gland region and evolve the gland region from the lumen. Hence, we refer to these methods as lumen-based methods (see Table 3.3). As a result, these methods suffer from the following limitations.

1. The tissue image is obtained from a 2D tissue slice sectioned from the prostate biopsy. Depending on the position in the biopsy that the tissue slice is sectioned, the lumen can be occluded or it may appear as multiple lumina in the 2D gland structure in the tissue image. Hence, segmenting a gland based on a particular lumen may create unexpected results, i.e., no detection of glands without lumen or multiple segments detected for one gland. Figure 3.6 illustrates these segmentation issues.

2. For tissue image classification purposes, not being able to detect glands without lumen can affect the Gleason grade 3 vs Gleason grade 4 classification result. In many grade 4 images and some of the grade 3 images, lumina are commonly occluded, leading to fewer glands being detected from the entire images. Hence, a certain amount of information (from the non-detected gland regions) is not utilized when computing features to represent these images (see Figure 3.7).

(a)   (b)
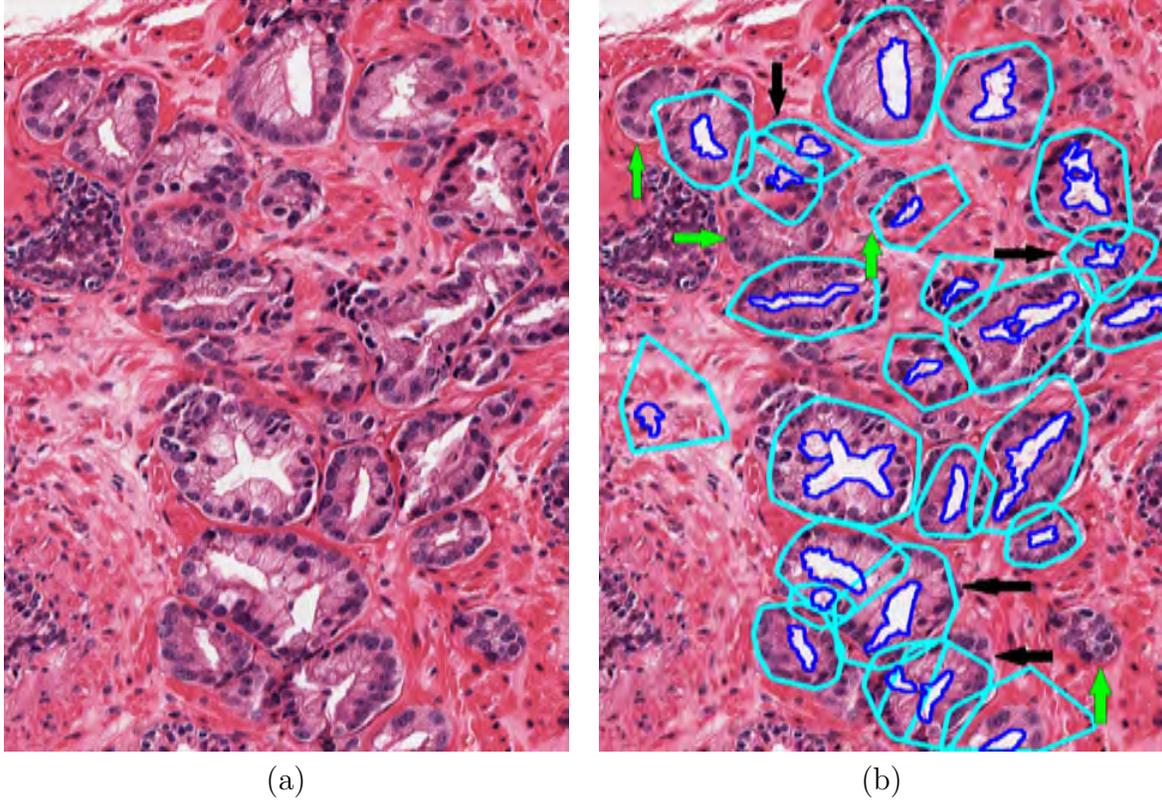
Figure 3.6: The gland segmentation result of the NLA method (a lumen-based method) on a grade 3 tissue image. (a) A grade 3 image of $1,100 \times 750$ pixels (corresponding to a 0.55×0.37 mm^2 tissue region digitized at $20\times$ magnification). (b) The segmentation result of the NLA method, in which (i) multiple gland segments are detected for a single gland since multiple lumina are detected within the gland (indicated by black arrows), and (ii) glands without lumen are not detected (indicated by green arrows). Detected lumina are shown as blue contours.

In this section, we look at the gland segmentation problem from a different perspective (see Figure 3.8 for the illustration):

1. Each gland is considered as a group of epithelial nuclei (green rectangle)[4] that are close to each other, which may or may not contain lumen in the center (green and black arrows). Nuclei usually form a closed chain structure, or an ellipse on the gland boundary.

2. Stroma rarely appears inside the gland area but mostly in between different glands

---

[4]Epithelial nuclei are nuclei lying in the gland regions, as opposed to stromal nuclei that are nuclei scattered in the stroma region.

<div align="center">(a)   (b)</div>
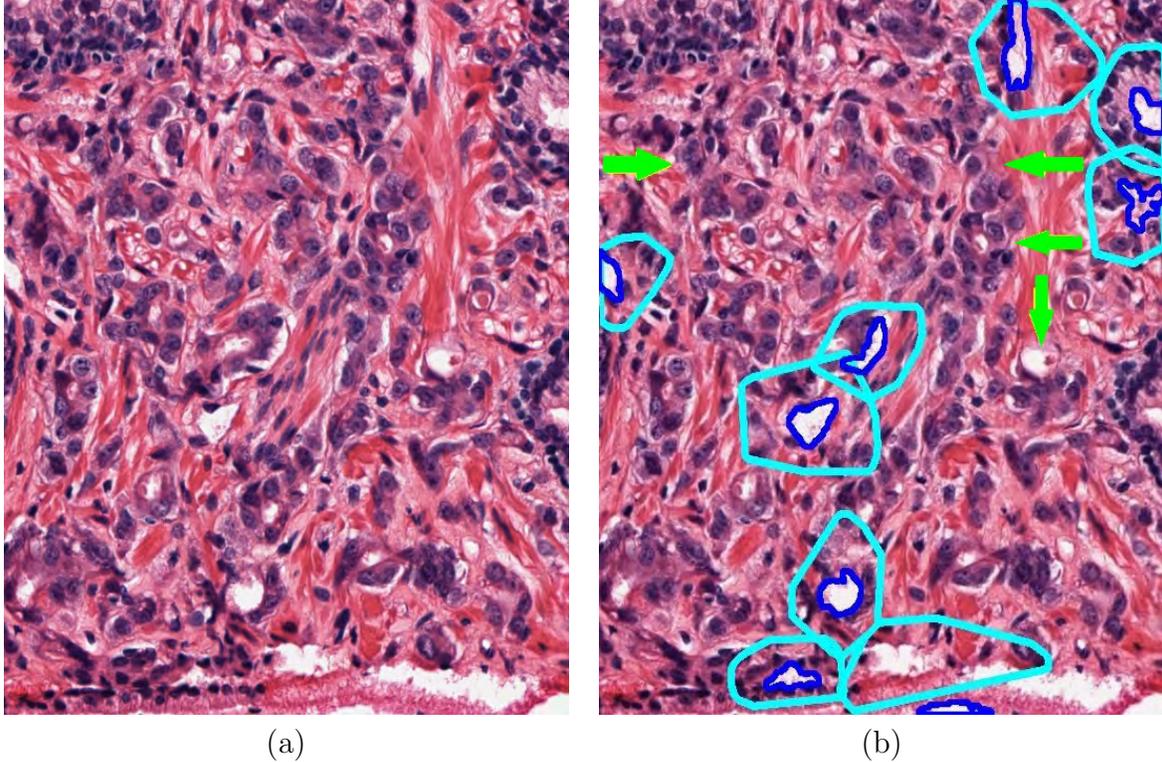
Figure 3.7: The gland segmentation result of the NLA method on a grade 4 tissue image. (a) A grade 4 image of 650×550 pixels (corresponding to a 0.32×0.27 mm^2 tissue region digitized at $20\times$ magnification). (b) The segmentation result of the NLA method, where only a few glands are detected from the entire image. The green arrows indicate examples of gland regions whose information is not used when computing features for classifying the image.

(the yellow contour). However, *glands are not perfectly separated by stroma*, i.e., we may see non-stroma area in between neighboring glands, or glands can be connected together (blue arrows).

Based on the perspective discussed above, we propose the following hypothesis. By mainly relying on nuclei when performing the segmentation (lumina are also used if available), we can come up with a gland segmentation method that overcomes the limitations of the lumen-based gland segmentation methods. Moreover, by using information about the nuclei arrangement in the segmented regions, we can improve the result of the grade 3 vs grade 4 tissue image classification. To prove this hypothesis, we develop a "nuclei-based" gland segmentation method and compute the "gland-scores" (described in the subsequent sections)

<div align="center">73</div>
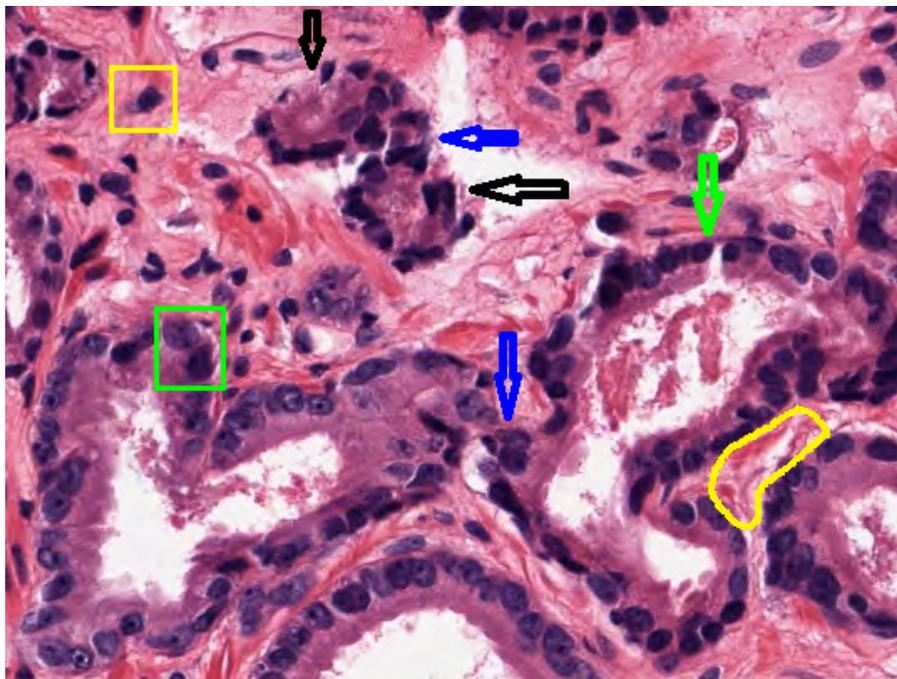
Figure 3.8: A different perspective for gland segmentation. A tissue image of 540×400 pixels (corresponding to a 0.27×0.2 mm^2 tissue region digitized at $20\times$ magnification), where the green rectangle denotes epithelial nuclei, the yellow rectangle denotes a stromal nucleus, the black arrows denote glands without lumen, the green arrow denotes a gland with lumen, the yellow contour denotes the stroma region between glands, and the blue arrows denote the areas where glands are connected (no intervening stroma).

of the segmented regions, which are then used to improve the grade 3 vs grade 4 tissue image classification result.

In the nuclei-based gland segmentation method, we aim at seeking a strategy to group the nuclei and lumina (if lumina are available) belonging to the same gland together. We achieve this goal by using graph theory techniques: we model the relationship between nuclei and lumina in the image by a nuclei-lumina-graph $\mathcal{G}$, where each nucleus or lumen is considered as a vertex in $\mathcal{G}$. Each edge in $\mathcal{G}$ is created as a "link" between a nucleus and a nucleus or between a nucleus and a lumen. The links indicate which nuclei and lumina are likely to belong to the same gland together. Finally, the normalized cut method [73] is applied on this graph to find and remove the weakest sets of links (the links with high potential to connect different glands) to partition them into different connected components, each of
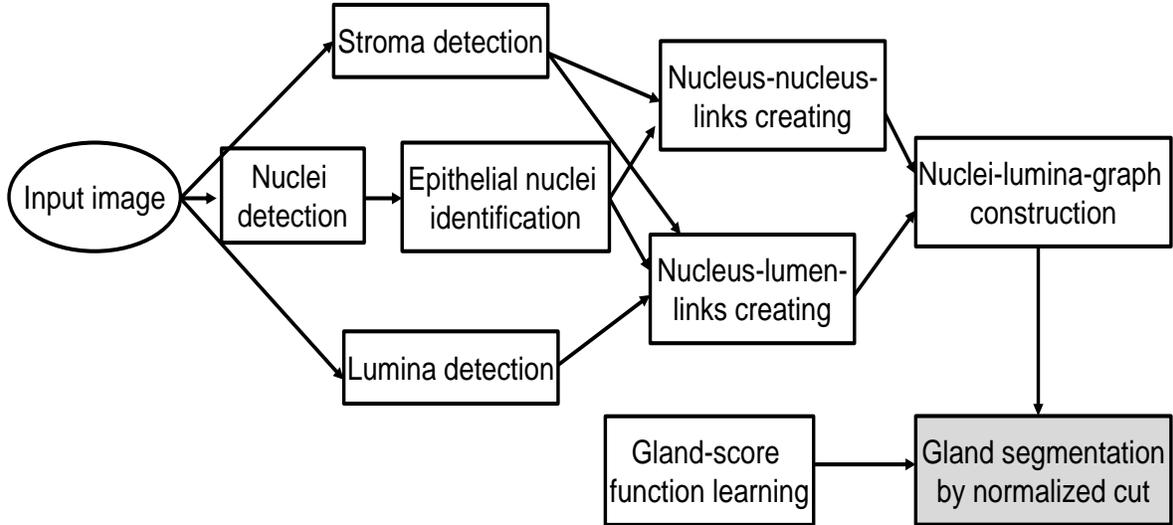
Figure 3.9: Flowchart of the nuclei-based gland segmentation method.

which corresponds to a gland (represented by a group of nuclei only or a group of both nuclei and lumina). A flowchart of the method is illustrated in Figure 3.9.

3.5.1 Nuclei Detection by Radial Symmetry

As mentioned in Figure 3.9, the first step of the method is to detect nuclei in the image. Since most nuclei have circular shapes and are of similar size, we use the radial-symmetry-based method [74] to detect the nuclei. This method is described in Algorithm 3.1 and Figure 3.10. The goal of the method is to detect the centers of the circular regions by using a voting scheme. To perform the voting, pixels with strong gradient magnitude in the image are selected (referred to as voting pixels) to cast the votes for its neighborhood region (voting region). Pixels with strong gradient magnitude (gradient magnitude is greater than a threshold τ_m) are chosen because they are likely to be the pixels on the nuclei boundary, which can effectively vote for the nuclei centers. The voting region, Ω_v, of a voting pixel p_v is a conical region with radius r_v, angle θ_v, and direction specified by the gradient vector at p_v (Figure 3.10d). The votes from all the voting pixels are accumulated to form a voting matrix V. Since the gradient vectors of pixels on the nuclei boundary point toward the nuclei center

(Figure 3.10c), the pixels in the nuclei centers receive the largest number of votes. Figure 3.10e shows the voting matrix (more reddish color indicates larger votes). Besides this voting matrix, we also estimate the nuclei radius in the matrix R using the voting distance. Next, we select the local maxima in this voting matrix (using the estimated nuclei radius) to use as the nuclei detection results. The advantage of this method is that it can detect the clumped nuclei (see the yellow arrow in Figure 3.10a). We choose the radius of the conical region as $r_v = 20$ pixels (the estimated maximum nuclei diameter) and the angle as $\theta_v = \pi/4$, so that this region is guaranteed to cover the nuclei center (Figure 3.10d).

Algorithm 3.1 Nuclei detection by radial symmetry

Input: Matrix of the gradient vectors in the image ∇I ($N \times M$), where $\nabla I(x, y)$ is the gradient vector at (x, y). Parameters τ_m, r_v, θ_v.

Output: Nuclei centers $\{n_i\}$

1: $V \leftarrow 0_{N,M}$ (voting matrix)
2: $R \leftarrow 0_{N,M}$ (nuclei radius)
3: $C \leftarrow 0_{N,M}$ (voting count)
4: **for each** $p_v = (x_v, y_v)$, where $\|\nabla I(x_v, y_v)\|_2 > \tau_m$ **do**
5:    $\Omega_v \leftarrow \{(x,y)|\|(x - x_v, y - y_v)\|_2 < r_v \text{ and } |\angle(x - x_v, y - y_v) - \angle\nabla I(x_v, y_v)| \leq \theta_v/2\}$
6:    **for each** $(x, y) \in \Omega_v$ **do**
7:      $d \leftarrow \|(x - x_v, y - y_v)\|_2$
8:      $V(x, y) \leftarrow V(x, y) + \dfrac{1}{d}$
9:      $R(x, y) \leftarrow R(x, y) + d$
10:     $C(x, y) \leftarrow C(x, y) + 1$
11:   **end for**
12: **end for**
13: $\forall x \in [1 \ldots N], y \in [1 \ldots M], R(x, y) \leftarrow R(x, y)/C(x, y)$
14: Using the maximum filter and the estimated radius in R to find the local maxima in V, i.e., nuclei centers $\{n_i\}$

3.5.2 Epithelial Nuclei Identification

As mentioned earlier, there are two types of nuclei in the image, namely epithelial nuclei (e-nuclei) and stromal nuclei (s-nuclei). E-nuclei are nuclei belonging to glands and are surrounded by cytoplasm, while s-nuclei are nuclei scattered in the stroma regions. In this step, we classify the detected nuclei into s-nuclei and e-nuclei. To perform the classification,
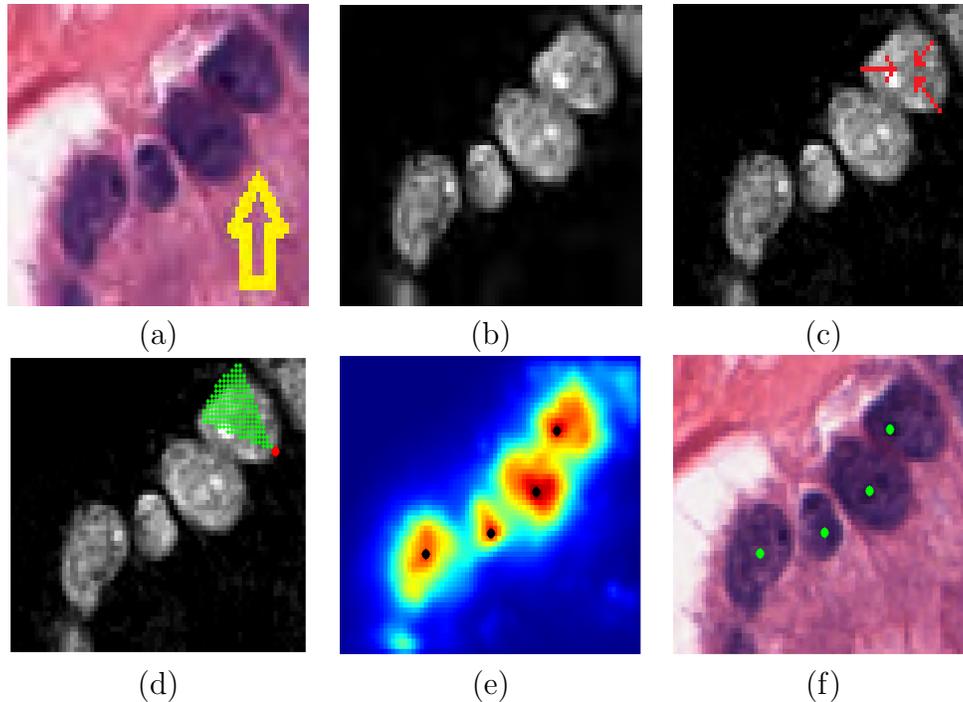
Figure 3.10: Nuclei detection by radial symmetry. (a) A tissue image of 58×74 pixels (corresponding to a 0.03×0.04 mm^2 tissue region digitized at $20\times$ magnification), where the yellow arrow indicates the two clumped nuclei. (b) The grayscale version of the image in (a). (c) The gradient vectors of the pixels on the nuclei boundary. (d) The voting region (denoted by green) of a voting pixel (denoted by red). (e) The voting matrix in which more reddish color indicates higher votes. The black dots indicate the selected local maxima in the matrix. (f) The final nuclei detection results (green dots) overlaid on the input image.

we compute textural features (Table 3.4) in the neighborhood of size S × S pixels (we use S = 40) of the nuclei centers detected from the previous step. These textural features capture the information about the neighborhood of the nuclei, which can be stroma regions (for s-nuclei) or cytoplasm regions (for e-nuclei). Hence, they can be used in discriminating the two nuclei types. A training set of e-nuclei and s-nuclei is obtained and a SVM classifier (RBF kernel) is trained for this classification task. Figure 3.11 shows an example of the classification result. The e-nuclei detected from the image are used for future processing (while the s-nuclei are discarded). For simplicity, we will use the term "nuclei" to indicate e-nuclei for the remainder of this section.
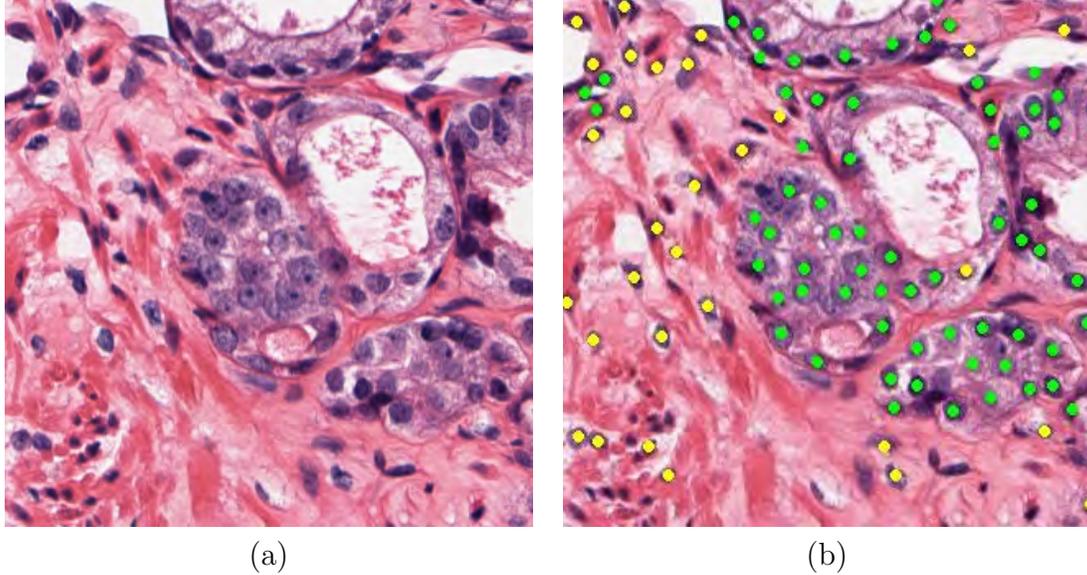
Figure 3.11: Nuclei classification result. (a) A tissue image of 320×320 pixels (corresponding to a 0.16×0.16 mm^2 tissue region digitized at $20\times$ magnification). (b) Nuclei classification result where yellow dots denote s-nuclei and green dots denote e-nuclei.

Feature type	Feature description
First-order statistics (14 features)	10-bin histogram, mean, standard deviation, median, and gradient magnitude of pixel intensity
Second-order statistics, i.e., features computed from the gray-level co-occurrence matrix (GLCM) (13 features)	Energy, correlation, inertia, entropy, inverse difference moment, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, and two information measures of correlation

Table 3.4: Textural features used for nuclei classification [4]. All the features are computed for all three channels of the *Lab* color space, resulting in 81 features in total.

3.5.3 Nuclei-Lumina-Graph Construction

To construct the nuclei-lumina-graph for the image as described in Figure 3.9, besides the nuclei, we need to also detect stroma and lumina from the image. Stroma and lumina are detected using the color-based clustering procedure described in section 2.3.3. The nuclei-lumina-graph demonstrates the relationship between nuclei and nuclei and between lumina and nuclei in the image, i.e., which nuclei and lumina should belong to the same gland together. We formalize the problem as following.

78

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the nuclei-lumina-graph to be constructed, where $\mathcal{V}$ denotes the vertex set and $\mathcal{E}$ denotes the edge set of the graph. Let $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$ and $\mathcal{L} = \{l_1, l_2, \ldots, l_L\}$ denote the sets of nuclei and lumina detected from the image, respectively. We construct $\mathcal{V}$ as $\mathcal{V} = \mathcal{N} \cup \mathcal{L}$. $\mathcal{V}$ can also be described as $\mathcal{V} = \{v_1^n, v_2^n, \ldots, v_N^n, v_1^l, v_2^l, \ldots, v_L^l\}$, where $\{v_1^n, v_2^n, \ldots, v_N^n\}$ are the N vertices corresponding to the nuclei set $\mathcal{N}$, and $\{v_1^l, v_2^l, \ldots, v_L^l\}$ are the L vertices corresponding to the lumina set $\mathcal{L}$. Our goal now is to construct the edge set $\mathcal{E}$ ($\mathcal{E} \leftarrow \emptyset$ at the beginning) to describe the relationship between the nuclei and lumina in the image. More precisely, if $\exists v_i v_j \in \mathcal{E}$, the two vertices v_i, v_j (v_i, v_j can be either a nucleus or a lumen) have potential to belong to the same gland.[5] To construct the edge set $\mathcal{E}$, we first develop two procedures, nucleus-nucleus-link creation and nucleus-lumen-link creation. Next, we discuss how to use these procedures in constructing $\mathcal{E}$.

3.5.3.1   Nucleus-Nucleus-Link Creation

We define a *link* between two nuclei n_i, n_j if they are likely to belong to the same gland. As mentioned earlier, our prior knowledge about nuclei and stroma in the image is that (i) nuclei of the same gland stay close together and form a closed chain, and (ii) stroma is more likely present between nuclei of different glands rather than between nuclei of the same gland. Based on this prior knowledge, we design an algorithm (Algorithm 3.2) to find all the nuclei that link to a nucleus of interest n_i. The algorithm is further illustrated in Figure 3.12.   The algorithm can be briefly described as follows. Given a nucleus of interest n_i, we consider the circular neighborhood Ω (with radius r_n) of n_i (the yellow circle in Figure 3.12b). Next, Ω is partitioned into k conical regions $\{\Omega_j\}$ (the conical regions separated by yellow lines in Figure 3.12b), each with an angle θ_n. At every Ω_j, we

1. Find the nucleus $n^* \in \Omega_j$ that is closest to n_i (the green stars in Figure 3.12c).

---

[5]We use the term "potential" since the final decision that the two vertices belong to the same gland or not will be decided by the normalized cut procedure applied on the graph (section 3.5.4).
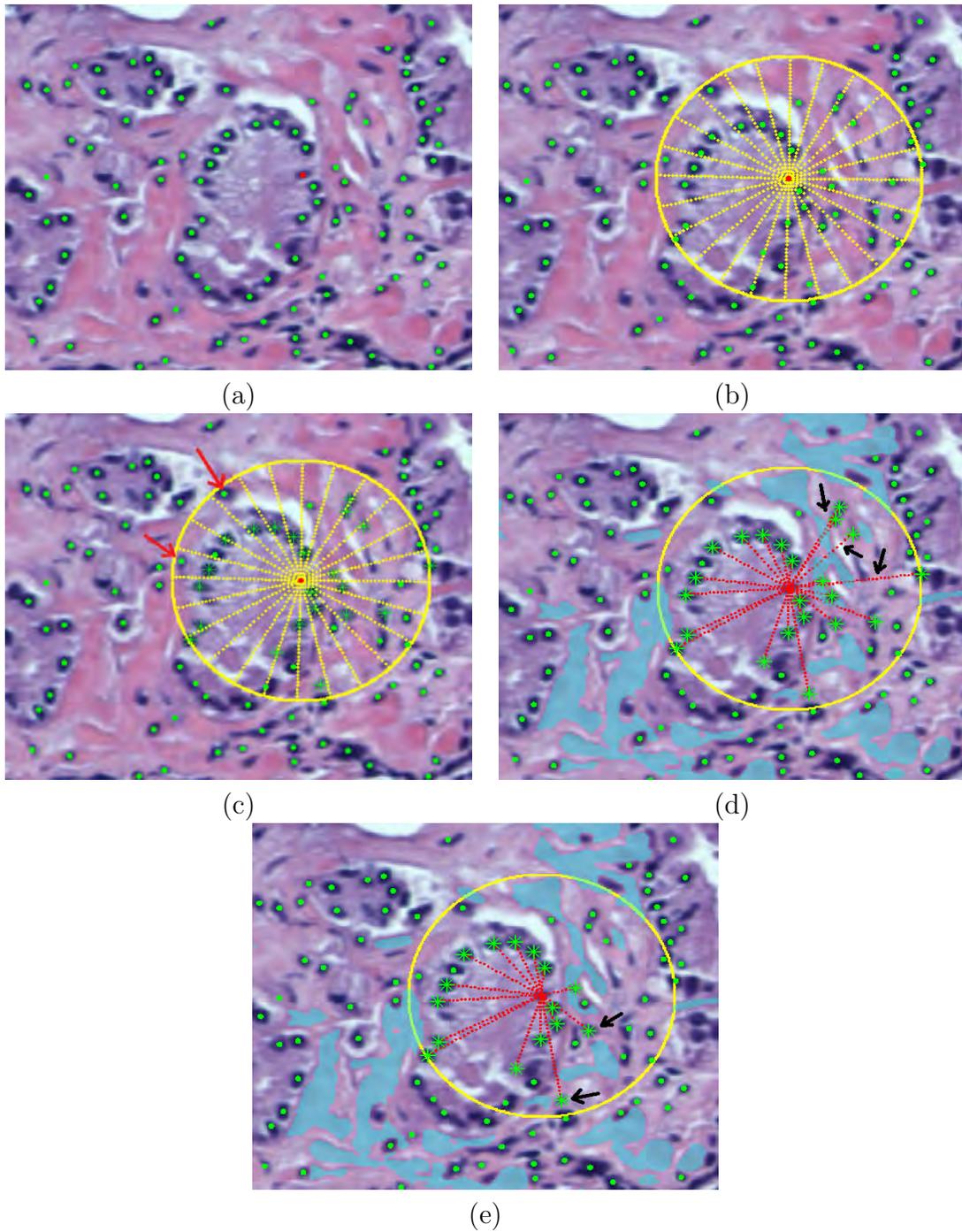
Figure 3.12: The nucleus-nucleus-link creation procedure. (a) A tissue image of 520×600 pixels (corresponding to a 0.26×0.3 mm^2 tissue region digitized at $20\times$ magnification), where the nucleus of interest n_i is indicated in red and the remaining nuclei are indicated in green. (b) The conical regions, shown in yellow. (c) The closest nucleus to n_i in each conical region, shown as a green star. (d) The detected stroma regions, shown in cyan and the lines connecting n_i to the closest nuclei, shown in red. The lines intersecting stroma are indicated by black arrows. (e) The final nuclei that link to n_i after discarding the nuclei with lines intersecting stroma, shown as green stars.

Algorithm 3.2 Nucleus-Nucleus-Link Creation

Input: Nuclei set $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$, with corresponding coordinates $\{(x_1^n, y_1^n), (x_2^n, y_2^n), \ldots, (x_N^n, y_N^n)\}$. Nucleus of interest $n_i = (x_i^n, y_i^n)$. Stroma mask $\mathbf{S}$. Parameters r_n, θ_n.

Output: The set of nuclei that link to n_i, denoted by Γ_i^n

1: $\Gamma_i^n \leftarrow \emptyset$
2: $\Theta \leftarrow [0, \theta_n, 2\theta_n, \ldots, 2\pi]$ (the list of angles with interval θ_n)
3: Generate $|\Theta| - 1$ conical regions:
4: $\forall j \in [1, |\Theta| - 1], \Omega_j = \{(x, y) | \|(x - x_i^n, y - y_i^n)\|_2 < r_n$ and $\angle(x - x_i^n, y - y_i^n) > \Theta_j$ and $\angle(x - x_i^n, y - y_i^n) < \Theta_{j+1}\}$
5: **for each** Ω_j **do**
6:     Find $n^* = (x^{n*}, y^{n*}) \in \Omega_j \cap \mathcal{N}$ such that $\|(x^{n*} - x_i^n, y^{n*} - y_i^n)\|_2 \leq \|(x^n - x_i^n, y^n - y_i^n)\|_2, \forall n = (x^n, y^n) \in \Omega_j \cap \mathcal{N}$
7:     Let $l(n^*, n_i)$ denote the line connecting n^* and n_i
8:     **if** $l(n^*, n_i) \cap \mathbf{S} = \emptyset$ **then**
9:         $\Gamma_i^n \leftarrow \Gamma_i^n \cup \{n^*\}$
10:    **end if**
11: **end for**

---

2. Examine the line $l(n^*, n_i)$ connecting n^* to n_i (the red lines in Figure 3.12d). If the line does not intersect the detected stroma area $\mathbf{S}$ (the cyan-highlighted regions in Figure 3.12d), we consider that n^* links to n_i (the green stars in Figure 3.12e).

Since the target of the nucleus-nucleus-links is to find glands without lumen, which are mostly small-sized and average-sized glands, we choose the radius $r_n = 100$ pixels, which corresponds to the size of an average-sized gland. For the conical angle θ_n, we select $\theta_n = \pi/12$ by estimating the density of nuclei on the boundary of the glands.

The reason for choosing the closest nucleus in each conical region is that this nucleus is most likely to belong to the same gland as n_i, as we can see there are some nuclei in Figure 3.12c that fall within the conical regions, yet do not belong to the same gland as n_i (e.g., the nuclei indicated by red arrows).

Although there are *bad links* (links between n_i and the nuclei not belonging to the same gland as n_i) created (e.g., the links indicated by black arrows in Figure 3.12e), these links are generally outnumbered by the *good links* (links between n_i and the nuclei belonging to the same gland as n_i), due to the assumption about stroma. In other words, the links

between nuclei within the same glands are denser than those between nuclei of different glands. Hence, by applying a global method like normalized cut (section 3.5.4), the bad links are likely to be removed to create the groups of nuclei that belong to the same gland.

3.5.3.2 Nucleus-Lumen-Link Creation

As has been mentioned earlier, the goal of the nucleus-nucleus-links is to deal with small-sized and average-sized glands without lumen. For glands with large size, lumen is commonly present. Hence, we utilize lumen to enhance the connection (the density of links) between nuclei within the gland. Given a lumen of interest l_i, we apply Algorithm 3.3 to find the nuclei that link to l_i. In this algorithm, we sample a number of points on the lumen boundary called lumen-points (similar to the NLA algorithm) and denoted as $\{p_j^l\}$. We apply Algorithm 3.2 on each point p_j^l to find the nuclei that link to p_j^l (i.e., we treat p_j^l as a nucleus of interest in Algorithm 3.2). The nuclei found are linked to the lumen l_i itself. By repeating the procedure for all the points in $\{p_j^l\}$, we obtain the nuclei set that link to l_i, denoted by Γ_i^l.

The radius r_l of the neighborhood region used at the lumen-point p_j^l is chosen as 50 pixels (the yellow circle in Figure 3.13b), which is the estimated maximum distance between lumen and nuclei on the gland boundary. The interval to sample the lumen-points on the lumen boundary is also chosen as r_l so that we can efficiently cover the region surrounding the lumen when finding nuclei.

Unlike the NLA algorithm that forms a gland by combining a lumen and the surrounding nuclei, in this algorithm, we only create the links between the lumen and the nuclei, but do not make the decision of forming the gland. Instead, we leave this decision to the normalized cut method discussed subsequently, which is able to make a global decision based on all the links in the image.
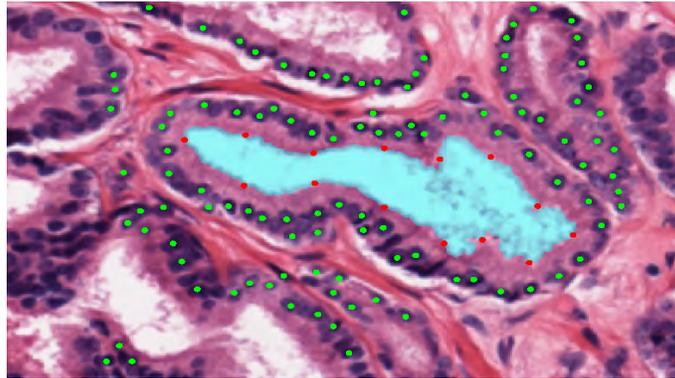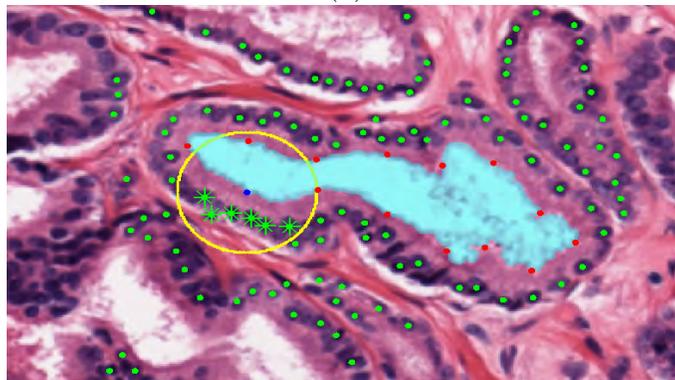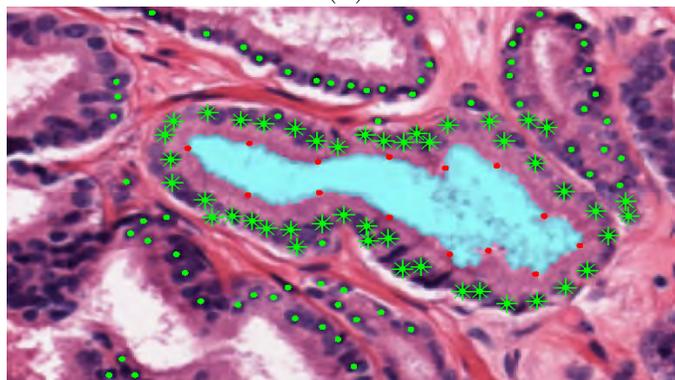
Figure 3.13: The nucleus-lumen-link creation procedure. (a) A tissue image of 520×790 pixels (corresponding to a 0.26×0.39 mm^2 tissue region digitized at $20\times$ magnification), where the cyan region denotes the lumen of interest, the red dots denote the lumen-points sampled on the lumen boundary, and the green dots denote the detected nuclei. (b) A selected lumen-point (blue dot) and nuclei that link to this lumen-point (green stars). (c) All the nuclei found (green stars) that link to the lumen.

Algorithm 3.3 Nucleus-Lumen-Link Creation

Input: Nuclei set $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$. Stroma mask $\mathbf{S}$. Lumen of interest l_i. Parameters r_l, θ_l.

Output: The set of nuclei that link to l_i, denoted by Γ_i^l

1: $\Gamma_i^l \leftarrow \emptyset$

2: Select $\{p_j^l\}$ by sampling at interval r_l from the boundary of l_i

3: **for each** p_j^l **do**

4:      Apply Algorithm 3.2 on p_j^l, with parameters (r_l, θ_l) to find the set of nuclei that link to p_j^l, denoted by Γ_p

5:      $\Gamma_i^l \leftarrow \Gamma_i^l \cup \Gamma_p$

6: **end for**

3.5.3.3 Constructing the Edge Set of the Nuclei-Lumina-Graph

By applying the two procedures, nucleus-nucleus-link creation and nucleus-lumen-link creation, on all the nuclei and lumina in the image, we are able to create all the links in the image. Note that, if two nuclei n_i and n_j have links to a lumen l_i, we also create a link for n_i and n_j, which is to strengthen the connection between nuclei in the same gland. We create an edge in the graph corresponding to each link in the image. Algorithm 3.4 demonstrates the procedure for constructing the edge set of the nuclei-lumina-graph.

Algorithm 3.4 Constructing the edge set $\mathcal{E}$ of the nuclei-lumina-graph $\mathcal{G}$

Input: Vertex set $\mathcal{V} = \{v_1^n, v_2^n, \ldots, v_N^n, v_1^l, v_2^l, \ldots, v_L^l\}$. Nuclei set $\mathcal{N} = \{n_1, n_2, \ldots, n_N\}$. Lumina set $\mathcal{L} = \{l_1, l_2, \ldots, l_L\}$. Stroma mask $\mathbf{S}$. Parameters $r_n, r_l, \theta_n, \theta_l$.

Output: The edge set $\mathcal{E}$

1: $\mathcal{E} \leftarrow \emptyset$

2: **for each** $n_i \in \mathcal{N}$ **do**

3:      Apply Algorithm 3.2 on n_i, with parameters (r_n, θ_n) to find nuclei that link to n_i. Denote the vertices in $\mathcal{V}$ corresponding to these nuclei as V_n

4:      $\mathcal{E} \leftarrow \mathcal{E} \cup v_i^n v, \forall v \in V_n$

5: **end for**

6: **for each** $l_i \in \mathcal{L}$ **do**

7:      Apply Algorithm 3.3 on l_i, with parameters (r_l, θ_l) to find nuclei that link to l_i. Denote the vertices in $\mathcal{V}$ corresponding to these nuclei as V_l

8:      $\mathcal{E} \leftarrow \mathcal{E} \cup v_i^l v, \forall v \in V_l$

9:      $\mathcal{E} \leftarrow \mathcal{E} \cup v_i v_j, \forall v_i, v_j \in V_l$

10: **end for**

3.5.4 Normalized Cut for Gland Segmentation

Remember that the nucleus-nucleus-links and nucleus-lumen-links created are based on the local information at the nuclei and lumina, without considering the global structure of the glands in the image. As a result, besides the good links, we may also obtain some bad links that connect nuclei of different glands together (when non-stroma regions are present in between the glands). Therefore, the nuclei-lumina-graph created for an image are likely to contain different connected components, each of which corresponds to a group of glands (the group may contain a single gland or multiple glands connected together, depending on the links created between the nuclei and lumina). Figure 3.14 shows an example of the nuclei-lumina-graph constructed for an image. To segment the glands, we need to find a way to partition each connected component in the graph into different smaller components such that each of them corresponds to a gland. The normalized cut method [73] is a suitable solution for this task. Normalized cut is a global method, which takes into account all the links in the image and finds the weakest set of links for the partitioning. Intuitively, the weakest set of links mostly contains the bad links since the bad links are less dense than the good ones, which makes the gland segmentation results appropriate.

In this method [73], given a graph $G = (V, E)$, the aim is to seek a minimum cut to partition the graph into two components, A and B. A cut is defined as the total weight of the edges to be removed to disconnect G, i.e.,

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v), \tag{3.1}$$

where $w(u, v)$ denotes the weight of the edge uv. The normalized cut is defined as [73]

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(B, A)}{assoc(B, V)}. \tag{3.2}$$

In this equation, $assoc(A, V) = \sum_{u \in A, v \in V} w(u, v)$ denotes the total connection from the vertices in A to all the vertices in G. A similar definition is used for $assoc(B, V)$.

To find the minimum cut, we first denote the binary assignment of the vertices into A
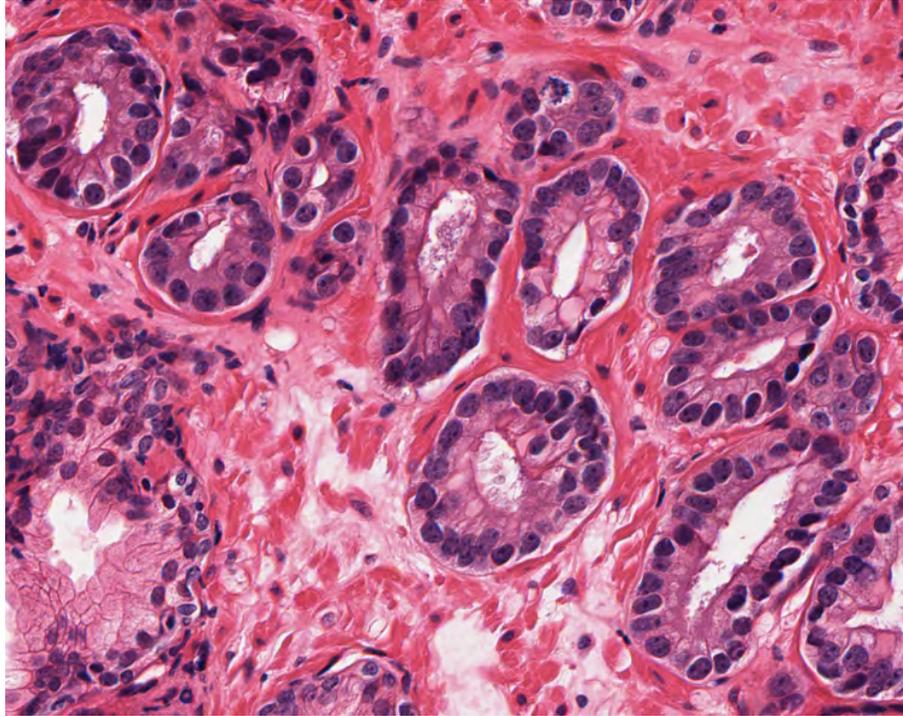
(a)
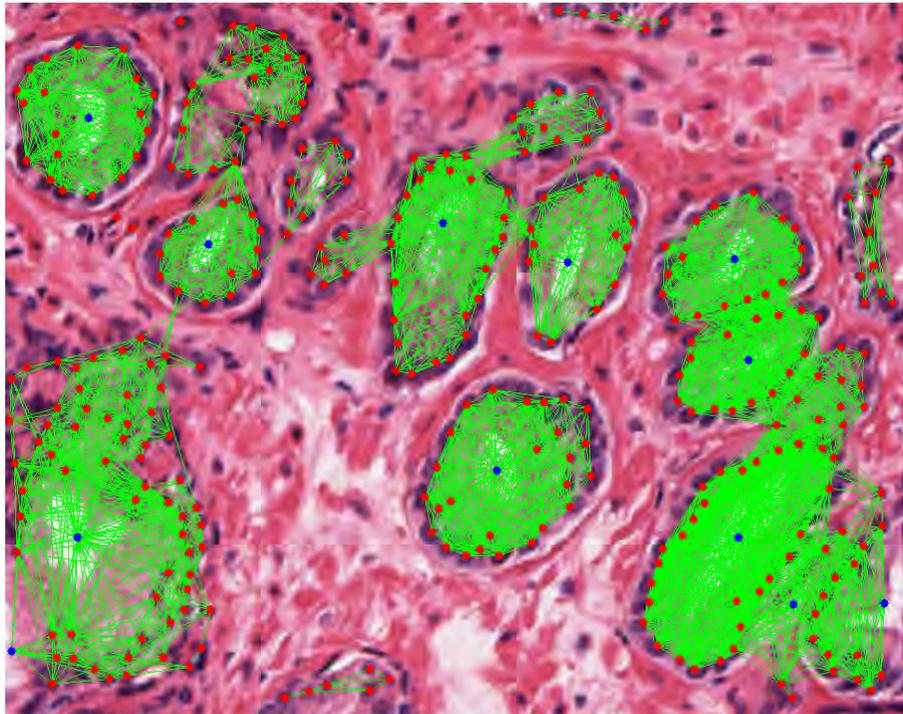
(b)

Figure 3.14: The nuclei-lumina-graph constructed for an image. (a) A tissue image of 750×560 pixels (corresponding to a 0.37×0.28 mm^2 tissue region digitized at $20\times$ magnification). (b) The nuclei-lumina-graph in which nuclei are denoted by red dots, lumina are denoted by blue dots, and the edges (links) are denoted by green lines.

and B as the binary vector $\mathbf{x} = \{x_1, x_2, \ldots, x_M\}$ ($M = |V|$), where $x_i = 1$ if the vertex v_i belongs to A and $x_i = -1$ if v_i belongs to B. Next, we denote $d_i = \sum_j w(i, j)$ as the total weight of the edges connecting the vertex v_i to all the other vertices in the graph. Using these notations, the *Ncut* formula in (3.2) can be rewritten as (we also denote $w(i, j)$ as w_{ij}):

$$
Ncut(A, B) = \frac{\sum_{(x_i > 0, x_j < 0)} -w_{ij} x_i x_j}{\sum_{x_i > 0} d_i} + \frac{\sum_{(x_i < 0, x_j > 0)} -w_{ij} x_i x_j}{\sum_{x_i < 0} d_i}. \tag{3.3}
$$

The goal now is to estimate $\mathbf{x}$ so that the *Ncut* value in (3.3) is minimized. By further defining an $M \times M$ diagonal matrix $\mathbf{D}$, with $\mathbf{D}(i, i) = d_i$, an $M \times M$ weight matrix $\mathbf{W}$ with $\mathbf{W}(i, j) = w_{ij}$, $k = \dfrac{\sum_{x_i > 0} d_i}{\sum_i d_i}$, $b = \dfrac{k}{1 - k}$, $\mathbf{y} = 1 + \mathbf{x} - b(\mathbf{1} - \mathbf{x})$, and performing further derivation as described in [73], the solution to $\mathbf{x}$ can be written as the following optimization problem

$$
min_{\mathbf{x}} Ncut(\mathbf{x}) = min_{\mathbf{y}} \frac{\mathbf{y}^T (\mathbf{D} - \mathbf{W}) \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}}. \tag{3.4}
$$

Instead of solving for $\mathbf{x}$, we now solve for $\mathbf{y}$. A possible solution to minimize (3.4) is to relax $\mathbf{y}$ to have real values and solve the eigenvalue system,

$$
(\mathbf{D} - \mathbf{W}) \mathbf{y} = \lambda \mathbf{D} \mathbf{y}. \tag{3.5}
$$

It was shown in [73] that the second smallest eigenvector of (3.5) is the solution for this normalized cut problem.

To apply the normalized cut method in our segmentation problem, we assign all edges in the nuclei-lumina-graph the same weight. More precisely, $\forall (v_i, v_j) \in \mathcal{V}$, $w_{ij} = 1$ if $v_i v_j \in \mathcal{E}$, otherwise $w_{ij} = 0$. Moreover, we perform the normalized cut in a recursive manner, i.e., we partition each connected component in our graph into two sub-components, and recursively partition the sub-components. One possible criterion that can be used to stop the recursive process is to examine the *Ncut* value and stop the cut if this value is higher than a predefined threshold δ_c (this means the connection within the graph is sufficiently solid). More precisely, when applying a cut on a component C, if $Ncut < \delta_c$, the cut is valid and the recursion continues. Otherwise, the cut is invalid and C is considered as one of the final segmentation

results.

An example of the recursive normalized cut procedure is shown in Figure 3.15. In this example, we demonstrate the recursive cut process applied on a graph (Figure 3.15b) created by a set of nuclei (green dots) and a lumen (the green dot enclosed in a black square) in Figure 3.15a. These nuclei and lumen belong to two glands in the image (Figure 3.15a). Applying normalized cut on the graph, we obtain two components denoted by the green and yellow dots in Figure 3.15c. The *Ncut* value for this cut is 0.15. Next, the two components in Figure 3.15c are further partitioned to obtain the results in Figures 3.15d and 3.15e, with $Ncut = 0.68$ and 0.72, respectively. By choosing the cut threshold $\delta_c = 0.5$, the cuts in Figures 3.15d and 3.15e are considered invalid, which means the two components in Figure 3.15c are the final segmentation results. In Figure 3.15f, we show the convex hull of these segments.

Since the local information (the nuclei distribution and the presence of stroma) at every gland is different, the densities of the good links and bad links in each connected component vary. Moreover, we do not have any information to estimate these densities, which means we do not know how to choose a suitable cut threshold δ_c to determine the final segmentation results. To address this issue, instead of finding a threshold δ_c, we aim to find a method to select the best components obtained during the recursive cut process to use as the final segmentation results. We define the best components as the components that are most similar to the gland. Intuitively, the arrangement of nuclei in a gland is similar to a *closed chain structure*, or in some cases, also similar to an *ellipse*. Figure 3.15c illustrates this intuition. Hence, we aim to develop different measures to estimate the similarity between the components (nuclei-groups)[6] to these structures. Using these measures, we come up with a single number, which we called *gland-score*, to estimate how similar the arrangement of the nuclei-group is to the arrangement of nuclei in a gland.

---

[6]Each component is a group of nuclei, which may also include lumina. Since we use only nuclei in the group for the subsequent computations, we also refer to these components as *nuclei-groups* in the remainder of the section.
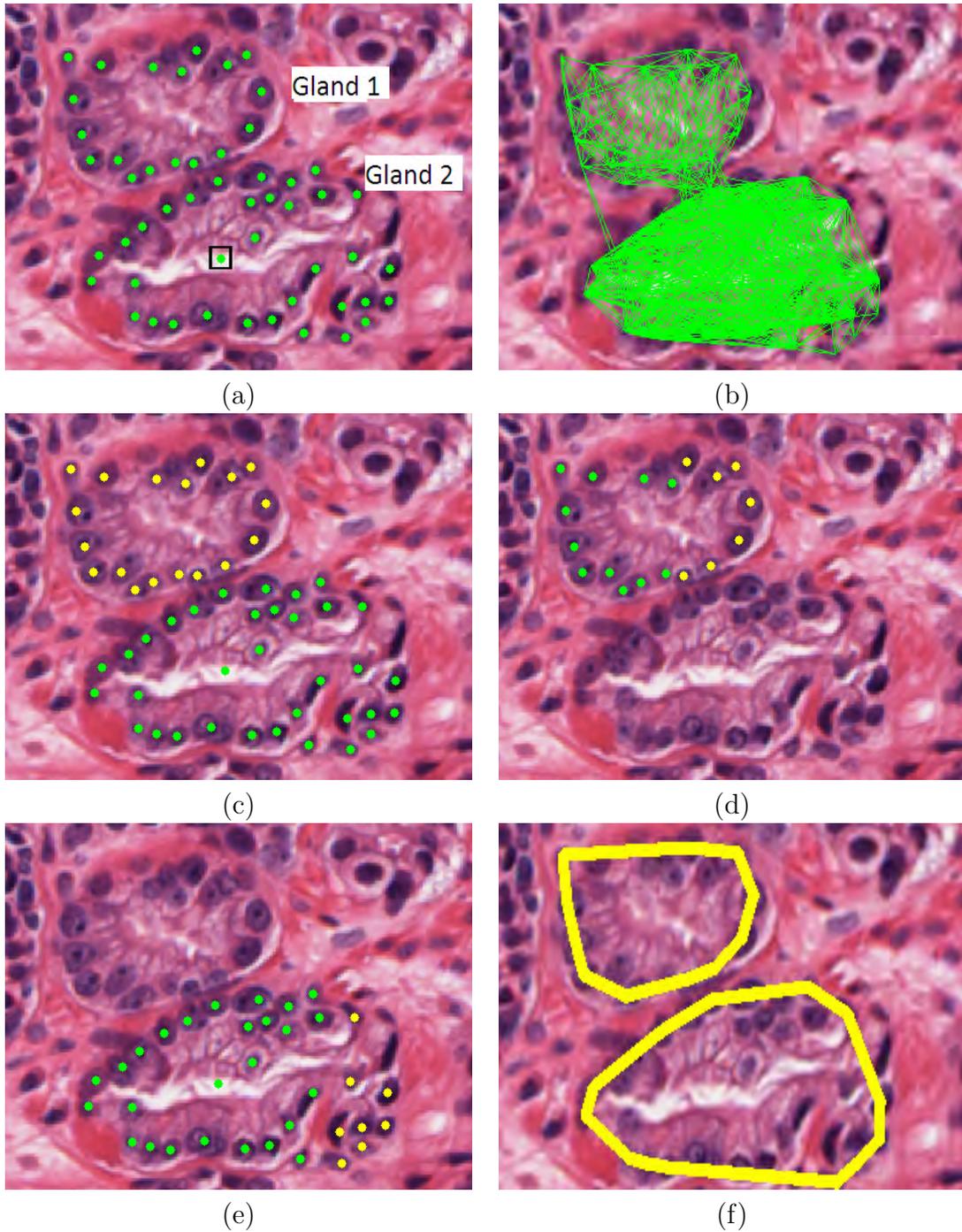
Figure 3.15: The recursive normalized cut process. (a) A tissue image of 330×440 pixels (corresponding to a 0.16×0.22 mm^2 tissue region digitized at $20\times$), in which the selected nuclei are denoted by green dots and the selected lumen is denoted by the green dot enclosed in a black square. (b) The nuclei-lumina-graph, whose edges are shown as green lines. (c) The result of the normalized cut applied on the graph in (b), where green and yellow denote the two components created ($Ncut = 0.15$). (d), (e) Results of the normalized cut applied on the two components in (c), with $Ncut = 0.68$ (d) and $Ncut = 0.72$ (e). (f) The final result showing two gland segments, when $\delta_c = 0.5$ is used.

For generalization purposes, *given a set of points $\{p_i\}$, we are interested in developing different measures to estimate how similar the arrangement of $\{p_i\}$ is to* (i) *a closed chain structure and* (ii) *an ellipse.*

3.5.4.1 Closed Chain Structure Measures

To compute the closed chain structure measures for a point set $\{p_i = (x_i, y_i)\}$, we first construct a graph $G = \{V, E\}$, where $V = \{p_i\}$ and $E = \{e = v_i v_j\} \ \forall v_i, v_j \in V$. The weight (or length) of the edge is computed as $w_{ij} = \|(x_i - x_j, y_i - y_j)\|_2$. We are interested in analyzing the structure of the minimum spanning tree (MST) of G [75, 76]. First, we denote the path between two vertices v_i, v_j in the MST as $P_{ij} = (v_i - v_j)$, where $|P_{ij}|_l$ denotes the length of the path (total length of the edges in the path), and $|P_{ij}|_v$ denotes the number of vertices in the path. We find the path with the largest number of vertices, P^*, i.e., $|P^*|_v \geq |P|_v, \ \forall P \in \text{MST}^7$, and refer to this path as the MST *backbone*. In Figure 3.16, we show the MSTs computed for the components obtained during the recursive cut of the graph mentioned in Figure 3.15. The MST backbones and the branches (the edges not belonging to the backbone) are shown as green lines and black lines, respectively, in this figure. We compute the following measures to estimate how similar the MST is to a closed chain structure (see Figure 3.16 for the intuition).

1. Mean degree (md): the average degree of non-leaf vertices (vertices with degree greater than 1) in the tree. So, $md \geq 2$, and a smaller md value indicates that the MST has fewer branches, which means it is more similar to the chain structure.

2. The ratio of the number of vertices in P^* to the total number of vertices, $r_v = \dfrac{|P^*|_v}{|V|}$. A larger value of r_v indicates that the MST has fewer branches or smaller branches, which means it is more similar to the chain structure.

---

[7]There may be several such paths but we select only one of them randomly.
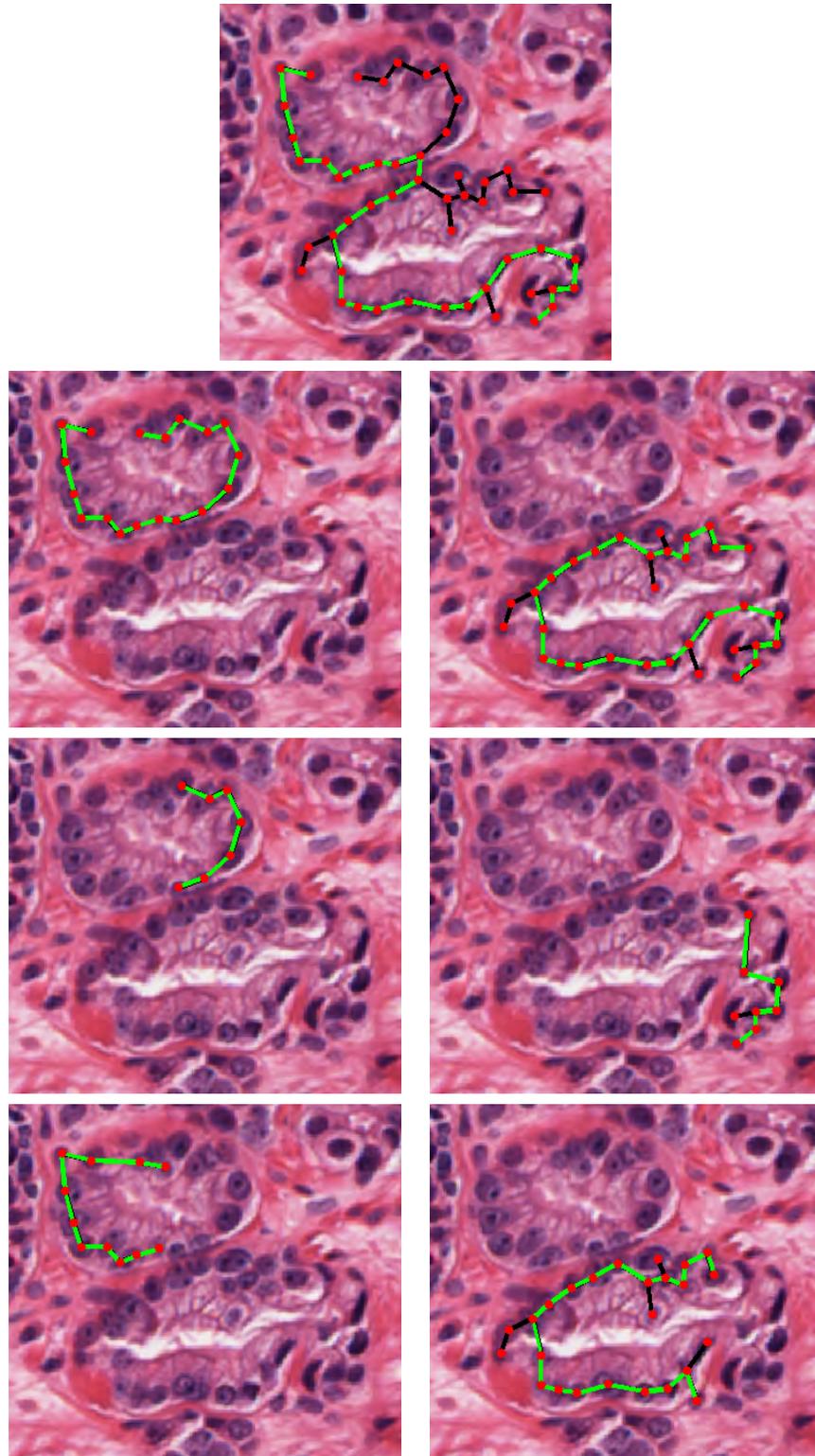
Figure 3.16: The MSTs and the MST backbones computed for the components obtained by the recursive cut process in Figure 3.15. The MST is shown as the green and black lines connecting the vertices (red dots). The green path is the MST backbone, while the black edges denote the branches in the MST.

3. The ratio of the length of P^* to the total length of all the edges in the MST, $r_l = \dfrac{|P^*|_l}{\sum_{e \in \text{MST}} |e|}$. Similarly, a larger value of r_l indicates that the MST is more similar to the chain structure.

4. The closeness index, ci, which determines the closeness of P^* (see Figure 3.17). To compute this measure, we first compute the center of P^*, $C_0 = (x_0, y_0)$, with $x_0 = mean(x_j)$ and $y_0 = mean(y_j)$, where $v_j = (x_j, y_j) \in P^*$. Next, we partition the region surrounding C_0 into b angular bins, $\{\Omega_i\}_{i=1}^{b}$, each with an angle of $\pi/12$ (which means $b = 24$). In Figure 3.17, the angular bins are separated by the black and yellow lines. We compute ci as the ratio of the number of bins that contain vertices in P^* to the total number of bins, i.e.,

$$ci = \frac{|\{\Omega_i | \exists v \in P^* \cap \Omega_i\}|}{b} \tag{3.6}$$

For example, in Figure 3.17, the bins not containing vertices in P^* are indicated by the black lines, while the bins containing vertices in P^* are indicated by the yellow lines on the left border (clockwise order) of the bins. A larger value of ci indicates that P^* is more similar to the closed path, e.g., the ci value of the path in Figure 3.17a is larger than that of the path in Figure 3.17b.

Non–densely-distributed points: In a closed chain structure, the points scatter on the chain, yet are not densely distributed. To describe this property, we create the Delaunay triangulation on $\{p_i\}$ and compute the mean μ_t and variance σ_t^2 of the length of the edges in the triangulation. If the points are densely distributed, μ_t is small. If the points are uniformly distributed, σ_t^2 is small. If the points scatter on the chain structure, both μ_t and σ_t^2 are large. In Figure 3.18, we give examples of two point sets (nuclei-groups) with similar number of points, one with points forming a chain structure (Figure 3.18a), and one with points densely distributed (Figure 3.18b). It is clear that the triangulation in Figure 3.18a contains longer edges than the one in Figure 3.18b.
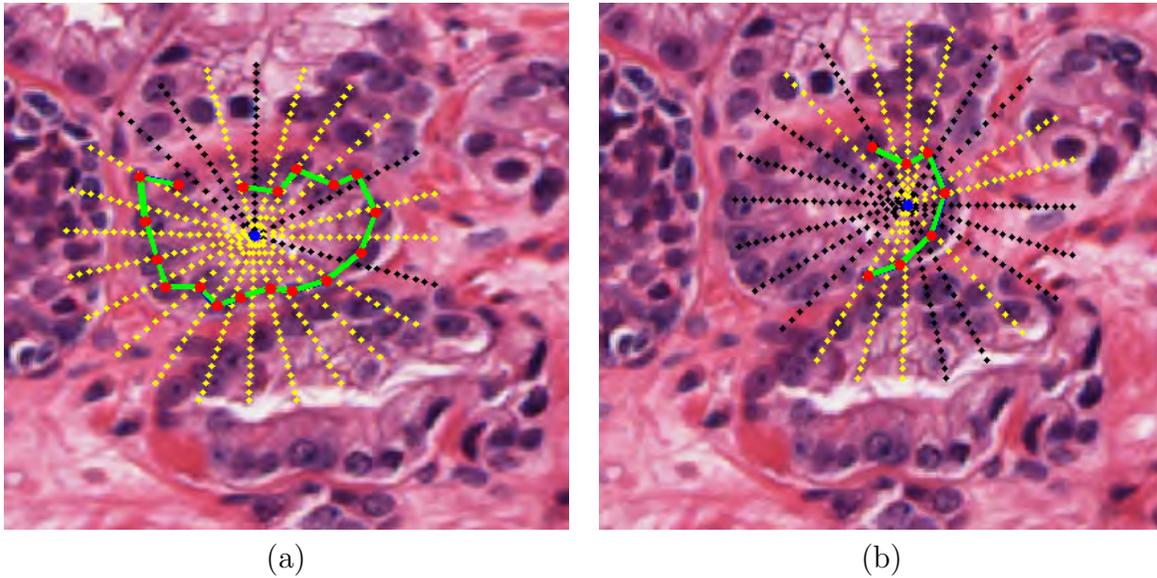
Figure 3.17: Computing the closeness index of a path. A path with a large value of closeness index (a) and a path with a small value of closeness index (b). The path is shown by the green line, whose vertices are shown as red dots. The center of the vertices is shown as the blue dot. The dotted lines separate the angular bins, where yellow lines denote the presence of vertices in the bin on its right (clockwise order), while the black lines denote the absence of vertices in the bin on its right (clockwise order).
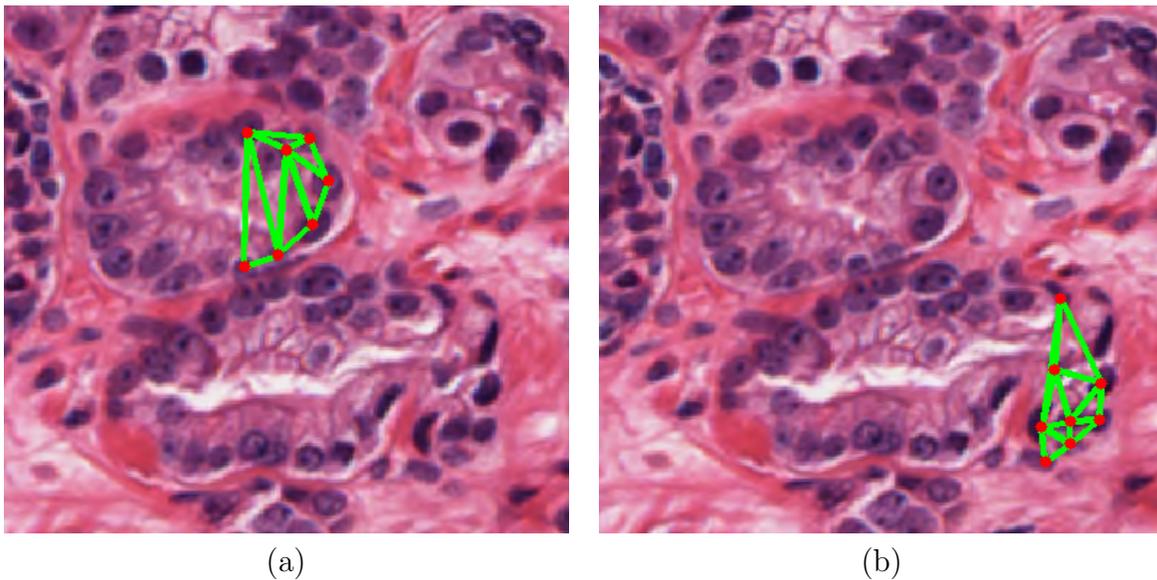
Figure 3.18: The Delaunay triangulation computed for the two components obtained from the recursive cut process in Figure 3.15. The nuclei in the component in (a) form a chain structure, while the nuclei in the component in (b) are more densely distributed.

3.5.4.2 Ellipse Measures

To estimate how similar the arrangement of a point set $\{p_i\}$ is to an ellipse, we first fit an ellipse to $\{p_i\}$. The conic equation of an ellipse is

$$E = ax^2 + bxy + cy^2 + dx + ey + f = 0. \tag{3.7}$$

We use the least square method to fit this ellipse model [77], i.e., estimate the parameters a, b, c, d, e using $\{p_i\}$. To make the fitting more robust to noise, the random sample consensus (RANSAC) algorithm [78] is employed. This RANSAC procedure assumes that the data contain both inliers (points that can be fitted to the model) and outliers (points that cannot be fitted to the model). RANSAC performs the following steps repeatedly for a predefined number of times T:

1. A sample subset is randomly selected from the original data and is used as the set of initial hypothetical inliers. A model $\mathcal{M}_i$ is fitted to this subset.

2. The entire data are tested again $\mathcal{M}_i$. All the data points that can be fitted to $\mathcal{M}_i$ (the fitting errors are smaller than a predefined threshold) are now considered hypothetical inliers. If a sufficient amount of hypothetical inliers are found at this step, the model is considered as a good model, and we go to the next step. Otherwise, we restart from step 1, i.e., no suitable model is found at this iteration.

3. $\mathcal{M}_i$ is now re-fitted using all of the hypothetical inliers found in step 2.

4. The total fitting error for $\mathcal{M}_i$ is estimated and $\mathcal{M}_i$ is considered the model found at this iteration.

The model with the smallest fitting error found within the T iterations is selected as the final fitting result $\mathcal{M}^*$. Figure 3.19 shows the ellipse fitting results to the components obtained from the recursive normalized cut process mentioned in Figure 3.15.

According to the RANSAC algorithm, when applying the ellipse fitting procedure on
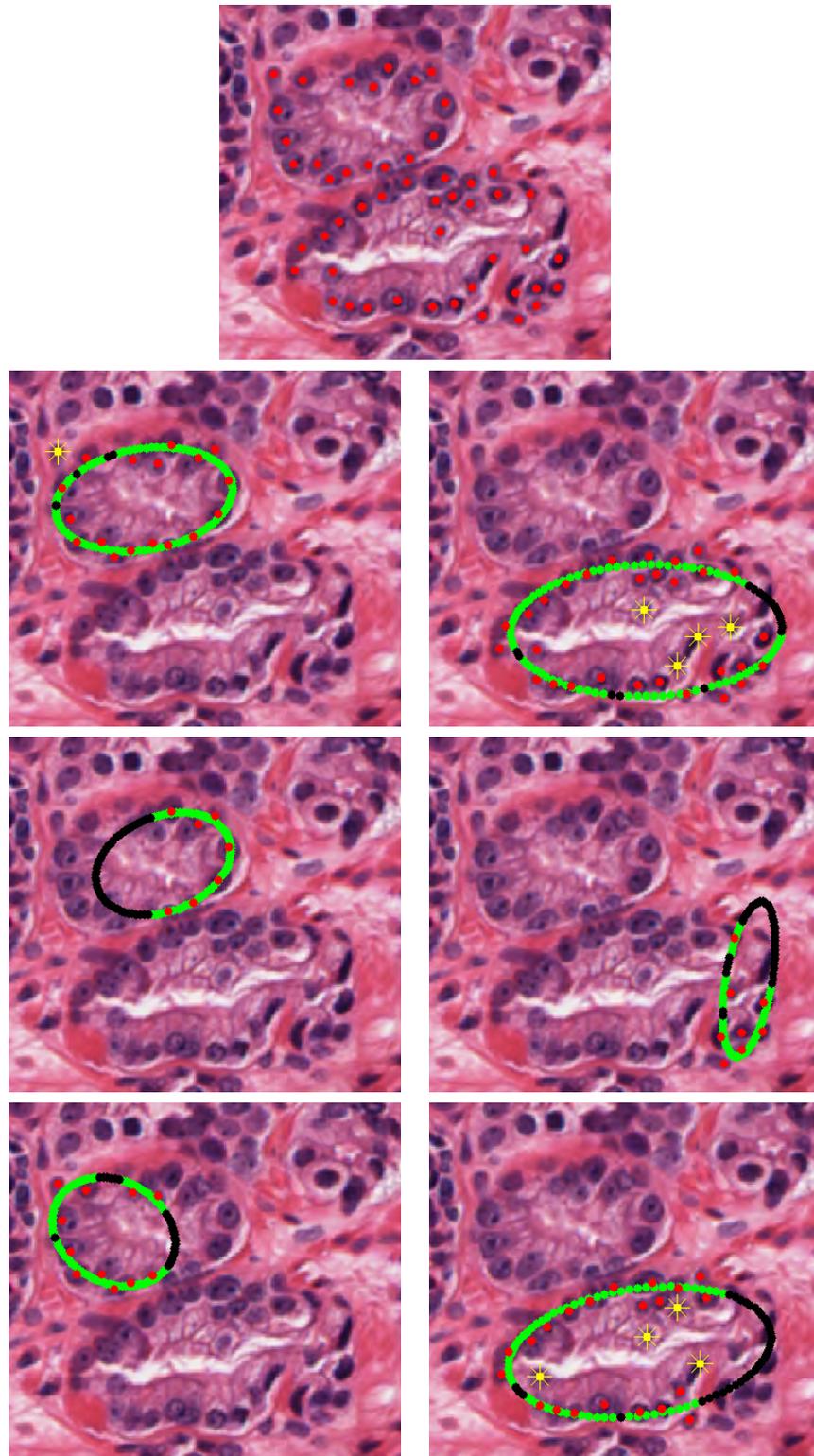
Figure 3.19: The ellipse fitting results for the components obtained by the recursive normalized cut process in Figure 3.15. An ellipse is not found in the original component. In the remaining components, the fitted ellipses are shown in green, while the fractions not covered by $\{p_i\}$ are shown in black. Inlier points are shown as red dots while outlier points are shown as yellow stars.

$\{p_i\}$, we may or may not find an ellipse. If an ellipse $\mathcal{M}^*$ is found, we compute the following measures[8].

1. The average fitting error. To compute this measure, we sample m points on the ellipse $\mathcal{M}^* = \{q_1, q_2, \ldots, q_m\}$. The fitting error for a point p_i is computed as $\epsilon_i = \min_{q_j} \|p_i - q_j\|_2$, $\forall q_j \in \mathcal{M}^*$. We average the errors for all the points in $\{p_i\}$. A smaller value of this measure indicates that $\{p_i\}$ is arranged more similarly to an ellipse.

2. The percentage of inliers, i.e., ratio of the number of inliers to the total number of points. A point $p_i \in \{p_i\}$ is considered an inlier if $\exists q^* \in \mathcal{M}^*$ such that $\|p_i - q^*\| < \delta_e$. We use $\delta_e = 13$ pixels (which is based on the estimated deviation of nuclei from the nuclei chain in the gland boundary). A larger value of this measure indicates that p_i is arranged more similarly to an ellipse.

3. The coverage index, i.e., the fraction of the ellipse that is covered by $\{p_i\}$. A point $q_i \in \mathcal{M}^*$ is considered covered by $\{p_i\}$ if $\exists p^* \in \{p_i\}$ so that $\|p^* - q_i\|_2 < \delta_e$. We compute the coverage index as the ratio of the number of points in $\mathcal{M}^*$ that are covered to the total number of points in $\mathcal{M}^*$. In Figure 3.19, the fraction of the ellipse that is not covered by $\{p_i\}$ is shown in black. A larger value of this measure indicates that p_i is arranged more similarly to an ellipse.

3.5.4.3   Computing the Gland-Score for a Nuclei-Group

We aim at computing all the closed chain structure measures and ellipse measures (which are summarized in Table 3.5) for a nuclei-group[9], and using them to derive a single number, called *gland-score*, to estimate how similar the arrangement of the nuclei-group is to the arrangement of nuclei in a gland. For convenience, we define the *gland-measure* vector

---

[8]If an ellipse is not found, we assign zero values to all these measures.

[9]The point set $\{p_i\}$ is the nuclei-group in this case. Recall that the components obtained from the normalized cut procedure can be considered nuclei-groups.

Type of measure	Measure description	Expected value when computed for a nuclei-group with high gland-score
Closed chain structure (need to build the MST and MST backbone)	Average degree of non-leaf vertices in the MST	low
	Ratio of the number of vertices on the MST backbone to the total number of vertices	high
	Ratio of the length of the MST backbone to the total length of all the edges in the MST	high
	Closeness index of the MST backbone	high
	Mean and variance of the length of the edges in the Delaunay triangulation	high
Ellipse (need to fit an ellipse)	Average fitting error	low
	Percentage of inliers	high
	Coverage index	high

Table 3.5: Summary of the gland-measures computed for a nuclei-group. There are nine measures in total.

of a nuclei-group C, denoted by $\mathbf{g}(C)$, as the nine-dimensional vector constructed by concatenating all measures in Table 3.5. In this table, we also mention the expected values of these measures when computed for a nuclei-group with a high gland-score. To compute the gland-score s of a nuclei-group C with gland-measure $\mathbf{g}(C)$, we need a function $\psi(\mathbf{g}(C)) = s$. We build this function using a learning framework: (i) obtain a training set of *gland nuclei-groups* (nuclei-groups corresponding to the complete glands) and *non-gland nuclei-groups* (nuclei-groups corresponding to parts of the gland, or nuclei-groups with random nuclei arrangement), (ii) compute the gland-measure vectors from them, and (iii) learn a classifier to separate the two groups, which is also served as the function ψ.

One possible way to obtain the training set is first performing the recursive normalized cut on the nuclei-lumina-graphs of the training images by using a certain value of the cut threshold δ_c as the stopping criterion. The resulting segments (nuclei-groups) are manually labeled as gland and non-gland segments to be used as the training data. For the training data, we use 130 non-gland segments and 100 gland segments, which contain a large variation

in shape and size, obtained from 30 different training images. The procedure is formalized in Algorithm 3.5. In this algorithm, we use SVM (RBF kernel) classifier, with probability output (the output of SVM is the probability that a nuclei-group is a gland) as the gland-score function ψ^{10}, which means $\psi \in [0, 1]$. The cut threshold δ_c is chosen as 0.5 for this training procedure (this value of δ_c produces the most appropriate gland segmentation results on the training images). Examples of the training data are shown in Figure 3.20.

Algorithm 3.5 Learning the gland-score function

Input: Training image set $\mathcal{I}$. Cut threshold δ_c.
Output: Gland-score function ψ, where $\psi(\mathbf{x}) \in [0, 1]$
1: $\Delta \leftarrow \emptyset$ (training nuclei-groups)
2: **for each** $I \in \mathcal{I}$ **do**
3:     Segment I using the recursive normalized cut procedure with δ_c as the stopping criterion, obtain segments (nuclei-groups) $\{C_i\}$
4:     $\Delta \leftarrow \Delta \cup \{C_i\}$
5: **end for**
6: **for each** $C_i \in \Delta$ **do**
7:     Manually assign a label y_i (gland or non-gland)
8:     Compute gland-measures, $\mathbf{g}(C_i)$
9: **end for**
10: Using $\{\mathbf{g}(C_i), y_i\}$ as the training data to learn a binary classifier ψ for gland vs non-gland classification, which produces probability output. ψ is used as the gland-score function

---

3.5.4.4   Using Gland-Score for Segmentation

We now discuss how to apply the gland-score function ψ in the recursive normalized cut process for gland segmentation. Basically, we compute the gland-score for each of the components created in this process and select the components with the highest scores. The algorithm is detailed in Algorithm 3.6. In this algorithm, we use $\delta_c = 1$, which is a very high value, so that we obtain a large number of segments during the process. As a result, we increase the chance of finding good segments.

    To explain Algorithm 3.6, we give an example in Figure 3.21[11], which re-uses the image

---

[10]We have also tried to use the kernel logistic regression method to estimate the probability output of the classification. However, the performance of this method is not as good as that of the SVM classifier. Hence, we use the SVM classifier for this task.

<center>(a)   (b)</center>
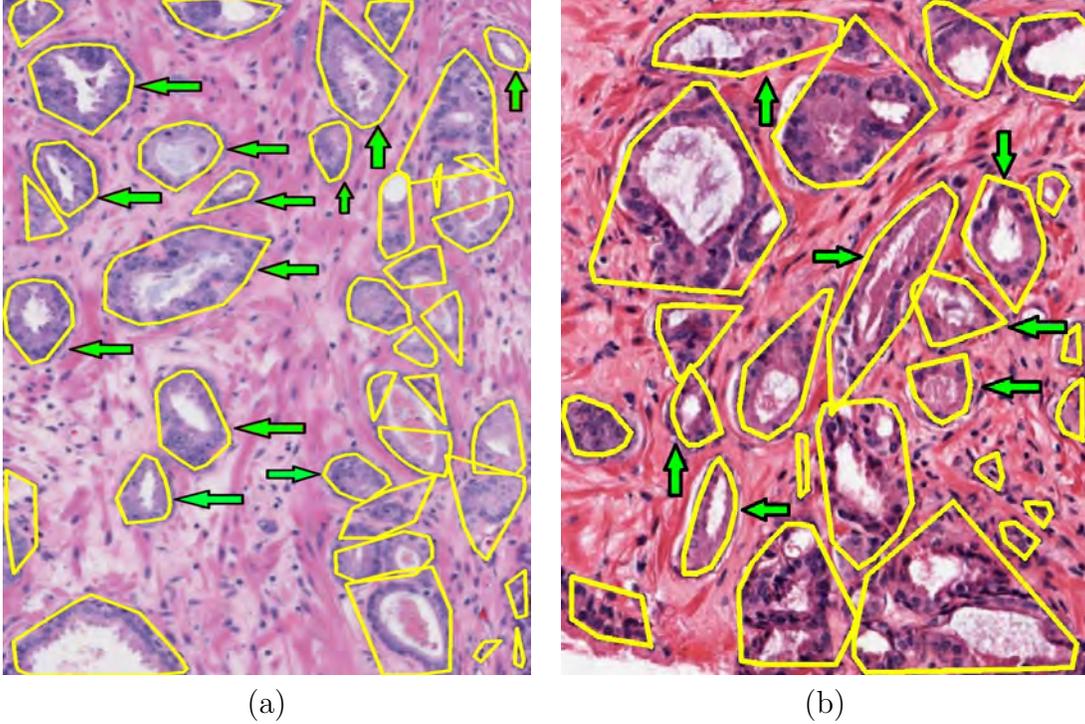
Figure 3.20: Examples of the training data used for learning the gland-score function. (a) A tissue image of $1,500 \times 940$ pixels (corresponding to a 0.75×0.47 mm^2 tissue region digitized at 20×). (b) A tissue image of 940×930 pixels (corresponding to a 0.47×0.47 mm^2 tissue region digitized at 20×). The segments indicated by green arrows are used as gland segments, while the remaining segments are used as non-gland segments in the training data. The cut threshold $\delta_c = 0.5$ is used for the recursive normalized cut segmentation in both the images.

in Figure 3.15a. In this example, we first perform the recursive normalized cut procedure (lines 3-10) to obtain the set of components $\Lambda_2 = \{C_0, C_1, C_2, C_{11}, C_{12}, C_{21}, C_{22}\}$ (note that these components are not mutually exclusive), and compute the gland-scores for these components (see Figure 3.21). Next, we sort the components based on the gland-scores (from high to low), which results in the order $C_2, C_1, C_{21}, C_{11}, C_{12}, C_0, C_{22}$ (the corresponding gland-scores are 0.96, 0.87, 0.77, 0.55, 0.17, 0.04, and 0.01, respectively). We iteratively choose the component with the highest score in Λ_2 (line 14) to include it into the final result Λ^*, such that it does not overlap the current components in Λ^* (line 16-17). As a result,

---

[11] When denoting the gland-score value s of a component C, instead of writing $\psi(\mathbf{g}(C)) = s$, we can simply write $\psi = s$.
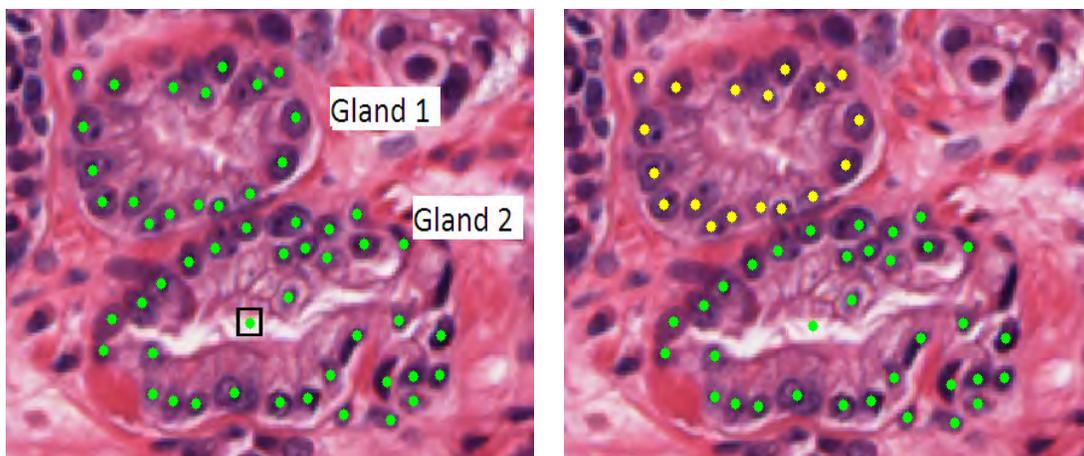
(a) Original component (C_0), $\psi = 0.04$

(b) Green component (C_1): $\psi = 0.87$, Yellow component(C_2): $\psi = 0.96$
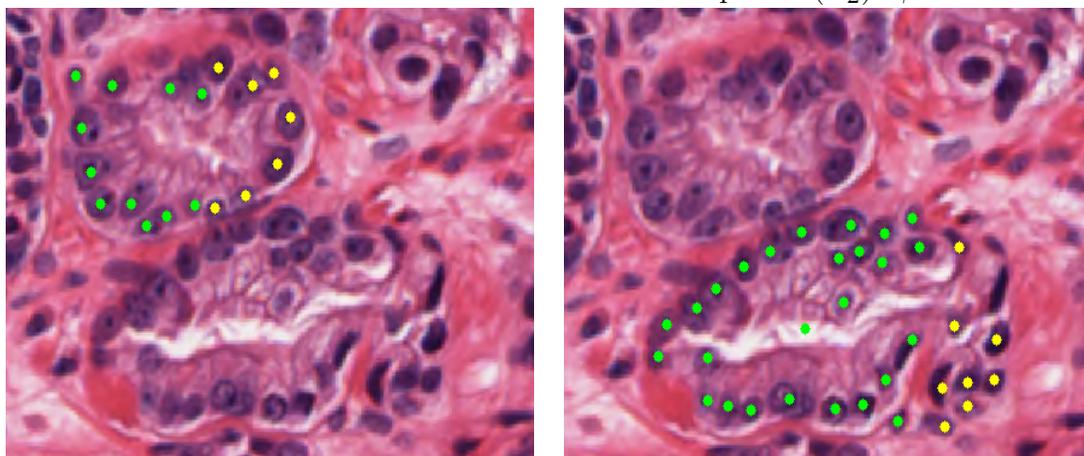
(c) Green component (C_{11}): $\psi = 0.55$, yellow component (C_{12}): $\psi = 0.17$

(d) Green component (C_{21}): $\psi = 0.77$, yellow component (C_{22}): $\psi = 0.01$

Figure 3.21: Computing gland-scores ψ for the components obtained by the recursive normalized cut process.

the components to be chosen are first C_2 and then C_1. We cannot choose other components because they all overlap C_2 or C_1. In general, by using this procedure, we obtain the final result $\Lambda^* = \{C_1, C_2, \ldots, C_k\}$ such that $C_i \cap C_2 \cap \cdots \cap C_k = \emptyset$ and $C_1 \cup C_2 \cup \cdots \cup C_k = C_0$, where C_0 denotes the original component.

Algorithm 3.6 Normalized cut gland segmentation using gland-score

Input: Nuclei-lumina-graph $\mathcal{G}$. Gland-score function ψ. Cut threshold δ_c.
Output: The set of components with the highest gland-scores, Λ^*
  1: $\Lambda_1 \leftarrow \mathcal{G}$
  2: $\Lambda_2 \leftarrow \emptyset$
  3: **while** $\Lambda_1 \neq \emptyset$ **do**
  4:     Select an arbitrary element $C \in \Lambda_1$
  5:     $\Lambda_1 \leftarrow \Lambda_1 \backslash \{C\}$
  6:     $\Lambda_2 \leftarrow \Lambda_2 \cup \{C\}$
  7:     Perform normalized cut on C to obtain C_1, C_2, and *Ncut* value
  8:     **if** *Ncut* $< \delta_c$ **then**
  9:         $\Lambda_1 \leftarrow \Lambda_1 \cup \{C_1, C_2\}$
 10:     **end if**
 11: **end while**
 12: $\Lambda^* \leftarrow \emptyset$
 13: **while** $\Lambda_2 \neq \emptyset$ **do**
 14:     Select $C^* = argmax_{C \in \Lambda_2} \psi(\mathbf{g}(C))$
 15:     $\Lambda_2 \leftarrow \Lambda_2 \backslash \{C^*\}$
 16:     **if** $\forall C \in \Lambda^*, C^* \cap C = \emptyset$ **then**
 17:         $\Lambda^* \leftarrow \Lambda^* \cup \{C^*\}$
 18:     **end if**
 19: **end while**

3.5.5   Qualitative Evaluation

We qualitatively compare the results of the nuclei-based gland segmentation method and the lumen-based gland segmentation method (the NLA method) in Figure 3.22. For the two images in this figure, the nuclei-based method is able to find glands without detected lumen, and not generate multiple segments for glands with multiple detected lumina as the lumen-based method.

3.5.6   Application to Grade 3 vs Grade 4 Tissue Image Classification

As reported in Table 3.2, we obtained a very good result for the normal vs cancer (grade 3 and grade 4) tissue image classification problem when using the lumen-based gland segmentation method and the structural-contextual features of the segmented regions. However, the result for grade 3 vs grade 4 classification is not as good. Hence, we are interested in using the
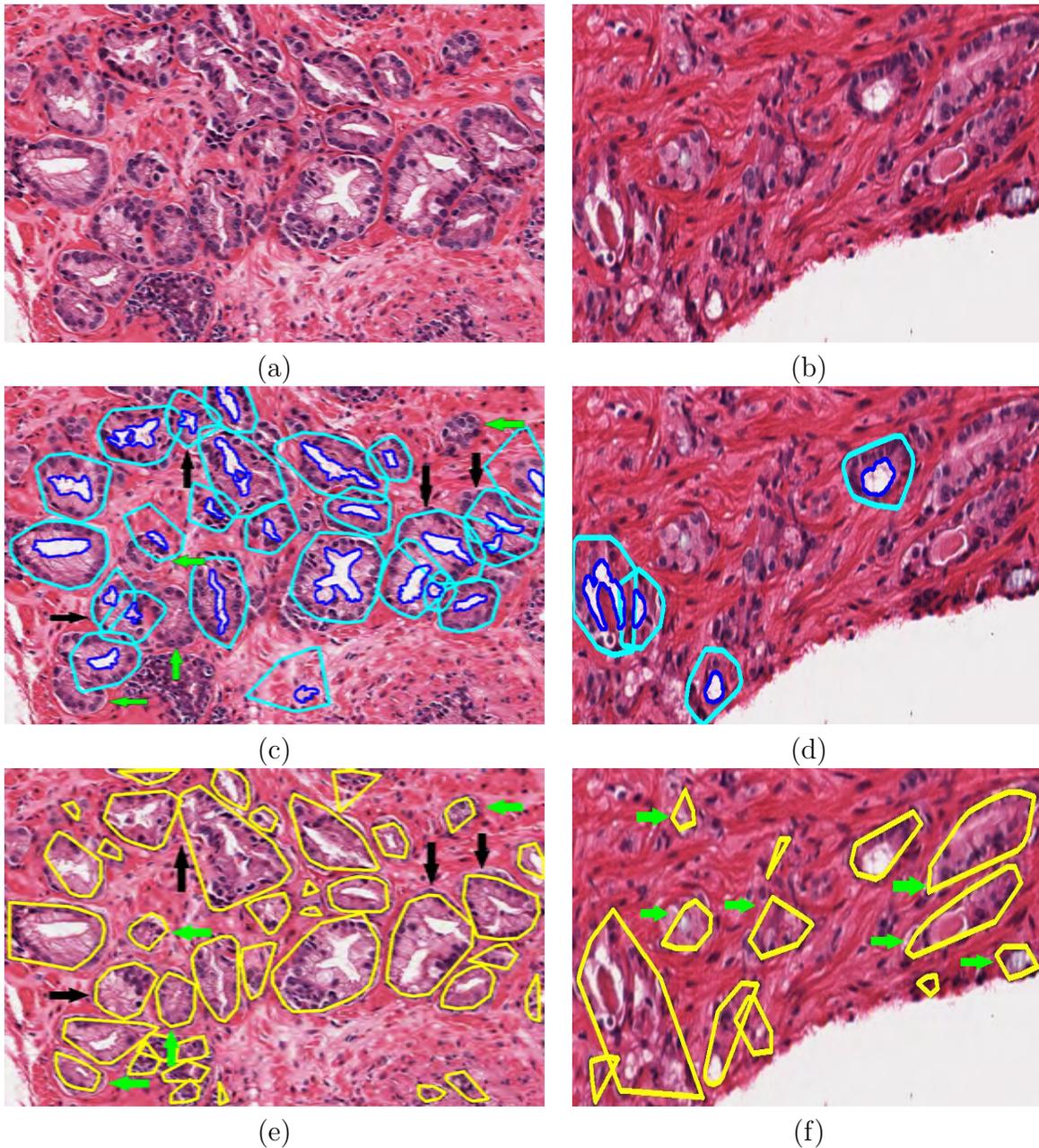
Figure 3.22: Comparison between the nuclei-based and the lumen-based gland segmentation methods. (a) A tissue image of $1,100 \times 770$ pixels (corresponding to a 0.55×0.38 mm^2 tissue region digitized at 20×). (b) A tissue image of 400×840 pixels (corresponding to a 0.2×0.42 mm^2 tissue region digitized at 20×). (c-d) The segmentation results of the lumen-based method, where detected lumina are shown as blue contours. (e-f) The segmentation results of the nuclei-based method. The black arrows in (c) and (e) indicate glands with multiple lumina, while the green arrows in (c), (e), and (f) indicate glands with no detected lumen. The nuclei-based method segments these glands successfully while the lumen-based method does not.

nuclei-based gland segmentation method and the gland-score function to improve the result of the grade 3 vs grade 4 tissue image classification problem.

3.5.6.1 Database

The database used in the following experiments includes all the Gleason grade 3 and grade 4 images mentioned in section 3.4. Recall that there are 134 grade 3 and 70 grade 4 images at 20× magnification (average image size is 1,200 × 900 pixels).

3.5.6.2 Computing Gland-Scores for the Segments in Grade 3 and Grade 4 Images

As mentioned in the Gleason grading method, in grade 4 images, gland structures degrade dramatically, i.e., we rarely see individual gland units, well-separated from each other, with nuclei arranging as closed chain structure or ellipse on the gland boundary. Instead, nuclei are likely to distribute randomly in the gland regions. As a result, when applying the nuclei-based gland segmentation method on grade 4 images, we are more likely to obtain non-gland segments, and less likely to obtain gland segments compared to grade 3 images. Figure 3.23 shows the comparison between the gland segmentation results of a grade 3 image and a grade 4 image.

Based on the observation mentioned above, we analyze the gland-measures and gland-scores of the segments in the two image types. In Figure 3.24, we show the MSTs and the fitted ellipses for the segments of the two images. In the grade 3 image, since there are several gland segments, the MSTs have fewer branches, the MST backbones are more similar to the closed chain structure, and the ellipses are better fitted to the segments than those in the grade 4 image, which contains mostly non-gland segments. We further compute the gland-scores for the segments in the two images, shown in Figure 3.25. It is reasonable that the gland-scores for the segments in the grade 3 image are much higher than those in the grade 4 image.
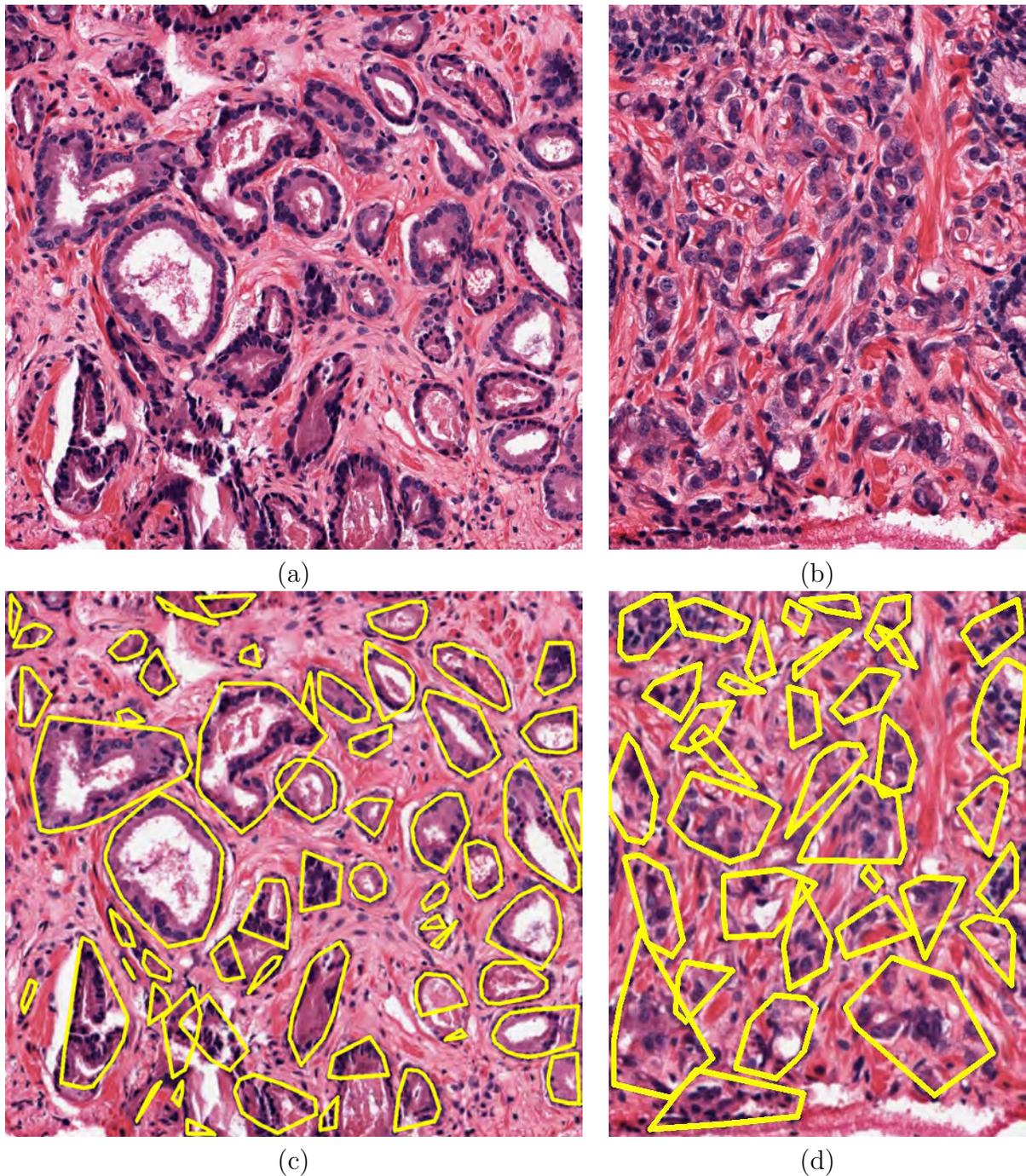
Figure 3.23: The nuclei-based gland segmentation results for a grade 3 and a grade 4 image. (a) A grade 3 image of $1,050 \times 940$ pixels (corresponding to a 0.52×0.47 mm^2 tissue region digitized at 20×). (b) A grade 4 image of 640×540 pixels (corresponding to a 0.32×0.27 mm^2 tissue region digitized at 20×). (c), (d) The segmentation results of the image in (a) and (b), respectively.
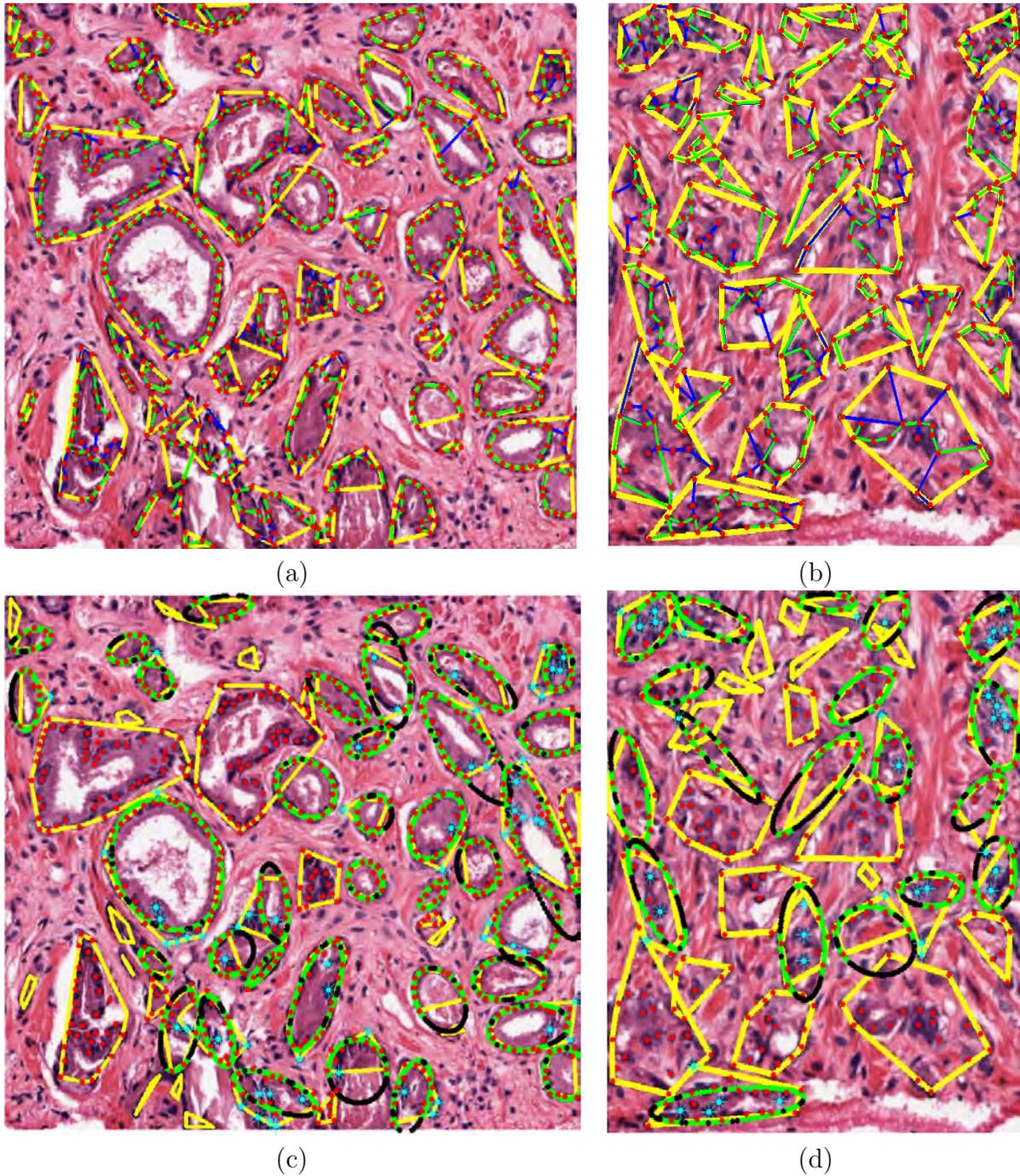
Figure 3.24: The MSTs and the fitted ellipses for the segments in the grade 3 and grade 4 images used in Figure 3.23. The nuclei are denoted by red dots. The MSTs are shown in (a) and (b), where the MST backbones are denoted by green lines and the branches are denoted by blue lines. The fitted ellipses are shown in green in (c) and (d), where the fractions of the ellipses not covered by nuclei are shown in black. The outlier nuclei are denoted by cyan stars.
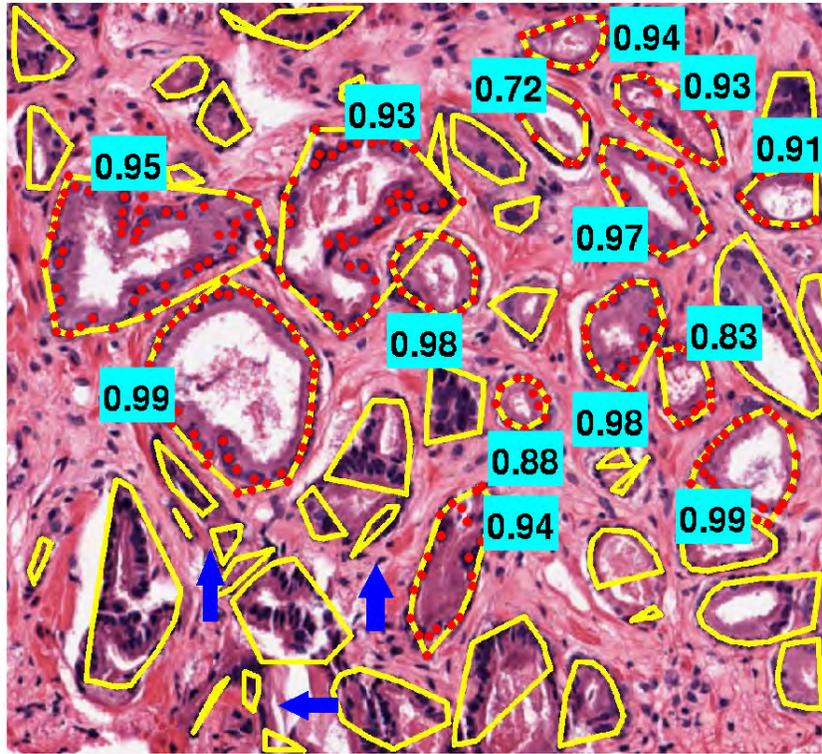
(a)
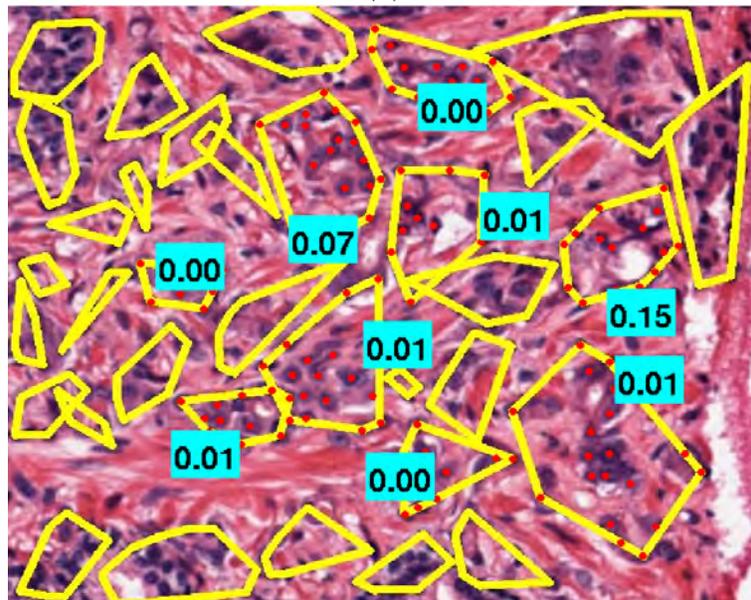
(b)

Figure 3.25: Gland-scores for some of the segments in the grade 3 and grade 4 images used in Figure 3.23. The selected segments are indicated by red dots (nuclei in the segments), with gland-scores showing in cyan boxes. The blue arrows in (a) indicate examples of the non-gland segments (noises) that will not be used in computing the image-gland-score. The gland-scores for the segments in the grade 3 image are higher than those in the grade 4 image.

3.5.6.3   Image-gland-score

To derive a single gland-score value to represent the image, referred to as *image-gland-score* and denoted as ψ^I, we compute the average gland-score of the segments in the image. Due to noises in the segmentation results (e.g., the non-gland segments indicated by blue arrows in Figure 3.25a), we use only the segments with the highest gland-scores to compute ψ^I. More formally, let $\psi_1, \psi_2, \ldots, \psi_n$ be the sorted gland-scores of n segments in the image, i.e., $\psi_1 > \psi_2 > \cdots > \psi_n$. We compute ψ^I as

$$\psi^I = \frac{\sum_{i=1}^{k} \psi_i}{k}, \text{where } k = \lceil \beta n \rceil (\beta \in [0, 1]). \tag{3.8}$$

So $\psi^I \in [0, 1]$ and β is the percentage of the number of segments that are used to compute ψ^I. Using this definition, we can compute ψ^I for all grade 3 images, and the mean of these values, denoted by $\mu(\psi_{g3}^I)$. Similarly, we compute ψ^I for all grade 4 images, and the mean of these values, denoted by $\mu(\psi_{g4}^I)$. As mentioned before, the gland-scores of the segments in grade 3 images are usually higher than those in grade 4 images, thus, we also have $\mu(\psi_{g3}^I) > \mu(\psi_{g4}^I)$. The difference between these two mean values, $d_\mu = \mu(\psi_{g3}^I) - \mu(\psi_{g4}^I)$, indicates the discrepancy of the ψ^I values of the grade 3 and grade 4 images. Since the value of d_μ depends on the percentage β, we plot the values of d_μ with respect to the values of β in Figure 3.26. The value $\beta = 0.2$ (i.e., only 20% of the segments with the highest ψ values in the image are used to compute ψ^I) results in the maximum value of d_μ, which is 0.24. At this value of β, we obtain $\mu(\psi_{g3}^I) = 0.72$ and $\mu(\psi_{g4}^I) = 0.48$.

By using $\beta = 0.2$, we plot the histogram of the ψ^I values for all grade 3 and grade 4 images in Figure 3.27. Although there is a certain overlap between the two distributions of the ψ^I values of the grade 3 and grade 4 images, we can still see the discrepancy between the two distributions. Hence, the ψ^I values could be helpful for the grade 3 vs grade 4 tissue image classification problem, as discussed in the next section.

In Figure 3.28, we show an example of a grade 4 image, which receives a very high ψ^I value ($\psi^I = 0.94$). In this image, the cancer is in the intermediate growth stage from grade
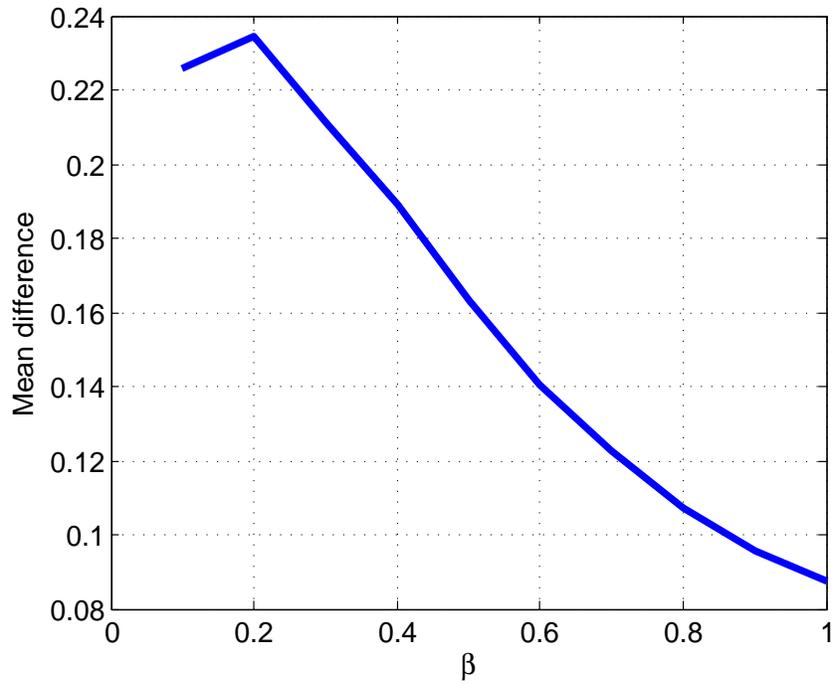
Figure 3.26: The plot showing the relationship between the percentage β and the mean difference $d_\mu = (\mu(\psi_{g3}^I) - \mu(\psi_{g4}^I))$.
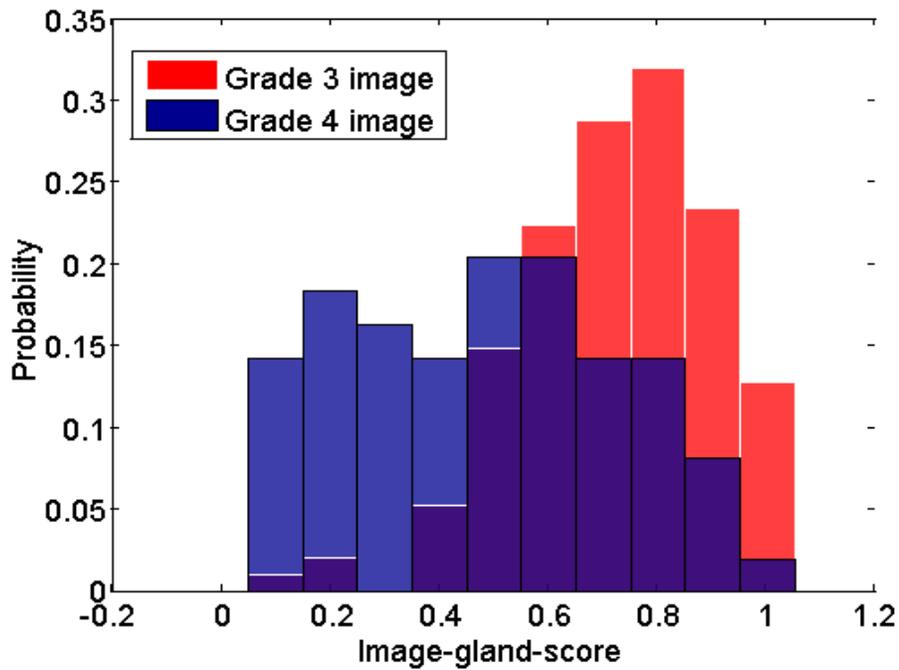
Figure 3.27: The distribution of the ψ^I values of grade 3 and grade 4 images computed using $\beta = 0.2$.
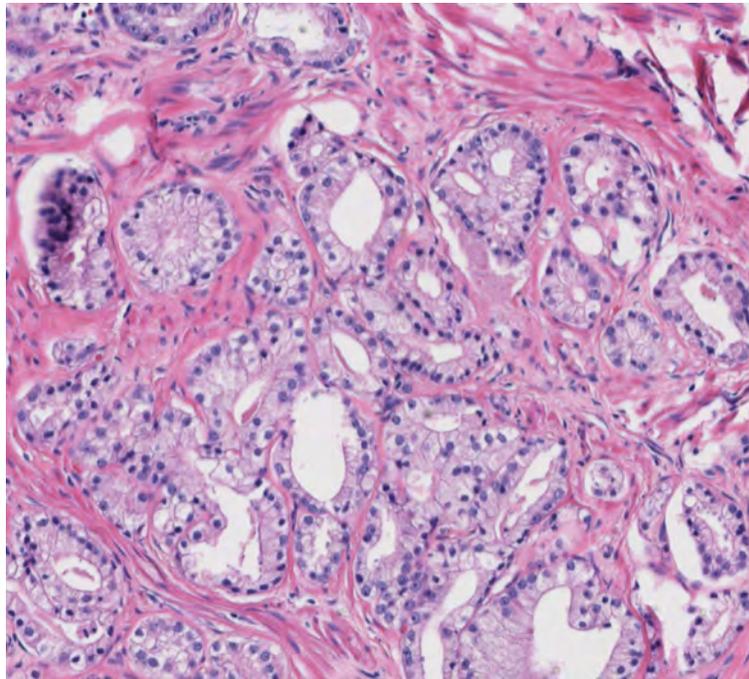
(a)
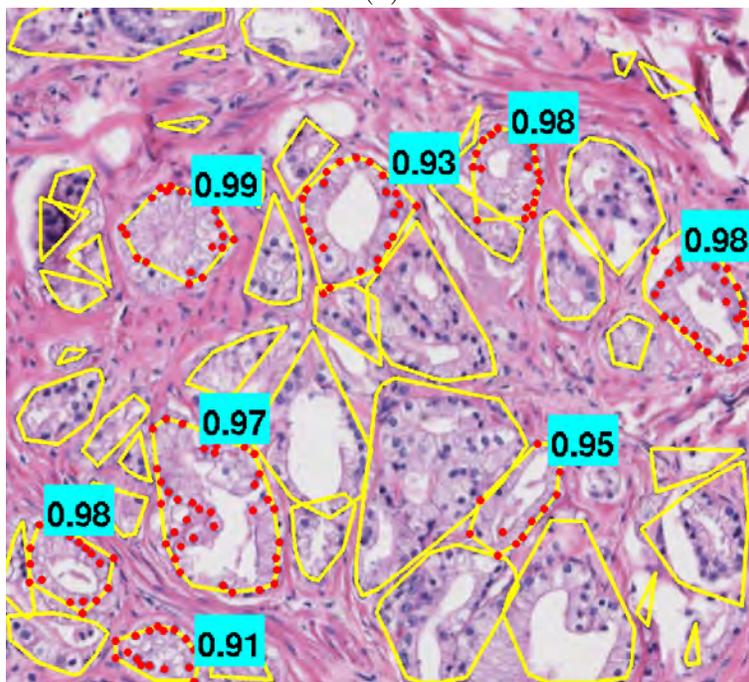
(b)

Figure 3.28: A grade 4 image with high ψ^I value. (a) A grade 4 image of $1,200 \times 1,300$ pixels (corresponding to a 0.6×0.65 mm^2 tissue region digitized at $20\times$). (b) The gland segmentation result of the image in (a), in which the segments with ψ values in the top 20% are indicated by red dots (nuclei in the segments). The ψ values of these segments are shown in cyan boxes, yielding $\psi^I = 0.94$.
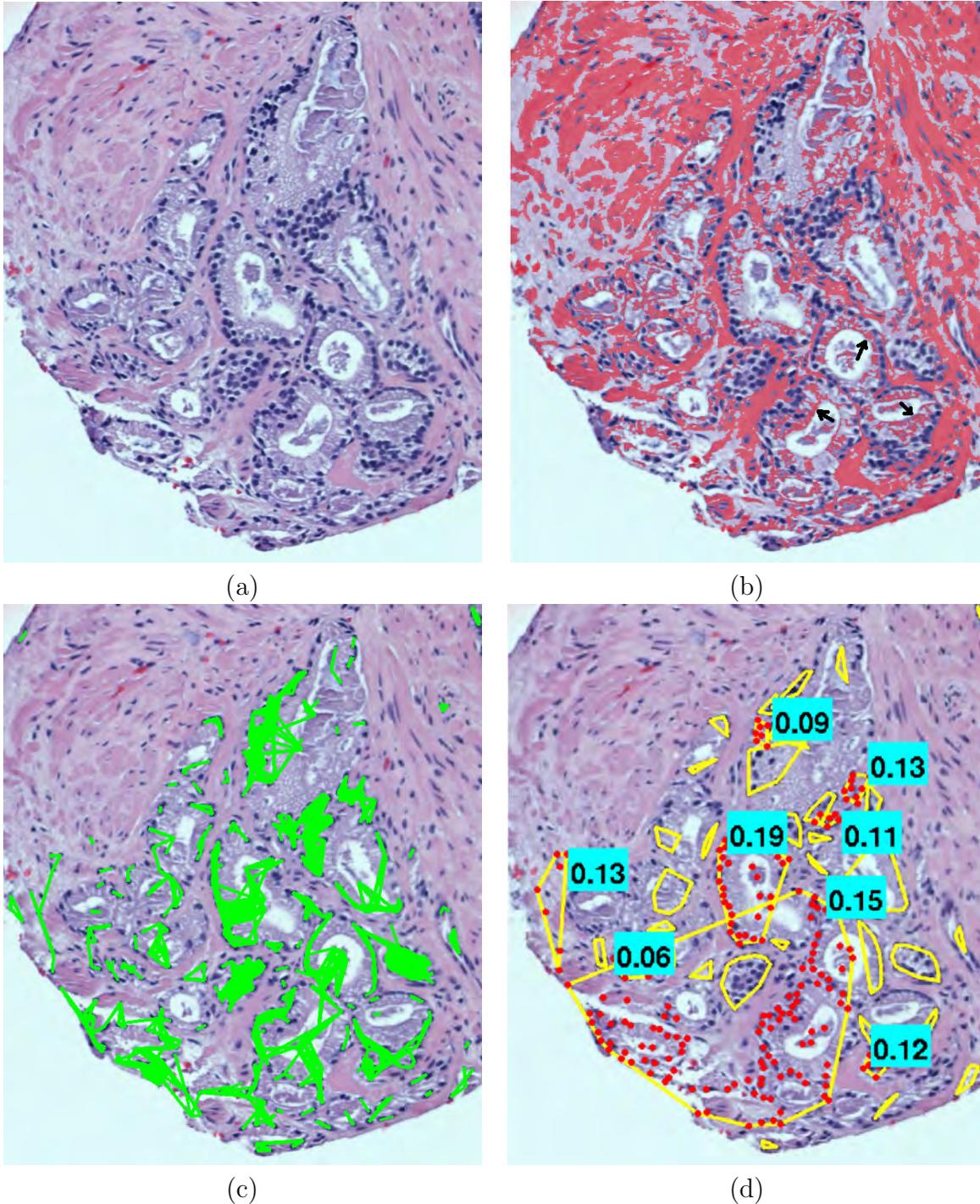
Figure 3.29: A grade 3 image with low ψ^I value. (a) A grade 3 image of $1,100 \times 940$ pixels (corresponding to a 0.55×0.47 mm^2 tissue region digitized at $20\times$). (b) The detected stroma regions in the image, shown in red. Rich stroma is mis-detected within the gland regions (indicated by black arrows). (c) The links created (denoted as green lines) showing the weak connection between nuclei within the same glands. (d) The gland segmentation result, in which the segments with ψ values in the top 20% are indicated by red dots (nuclei in the segments). The ψ values of these segments are shown in cyan boxes, yielding $\psi^I = 0.12$.

3 to grade 4. Hence, besides the fused glands, there are individual glands whose nuclei form closed chain structure. These individual glands are successfully segmented and obtain high ψ values, which explains the high ψ^I value of the image. In another example, a grade 3 image with a low ψ^I value is shown in Figure 3.29. This figure also shows the limitation of the nuclei-based segmentation method, the dependence on the detection of stroma in the image. Due to the false detection of rich stroma within the gland regions in the image (Figure 3.29b), the connection between nuclei of the same glands is weak (Figure 3.29c). Hence, the segments obtained do not capture the complete gland area but only some nuclei in the gland, thus, they receive low ψ values (Figure 3.29d).

3.5.6.4   A Fusion Method for Grade 3 vs Grade 4 Tissue Image Classification

As has been mentioned, a limitation of the nuclei-based gland segmentation method is that it can be affected by a poor stroma detection result. One possible solution to overcome this limitation when addressing the grade 3 vs grade 4 tissue image classification problem is to combine it with the NLA method. The NLA method does not rely on stroma to perform the gland segmentation (section 2.3), thus, it is not affected by the stroma detection results. Hence, we propose a fusion method to address this classification problem. In this fusion method, we use the following two methods to compute image features:

- Nuclei-based method: We segment glands using the nuclei-based gland segmentation method, compute the gland-scores of the segments, compute the image-gland-score ψ^I of the image, and use ψ^I as the image feature.

- Lumen-based method (the method used in the tissue image classification problem presented in section 3.3): We segment glands using the NLA method, compute 19 structural features of the gland segments (section 2.4)[12], and use the average of the features

---

[12]The contextual features are not computed because they are not useful to discriminate grade 3 and grade 4 tissue images.
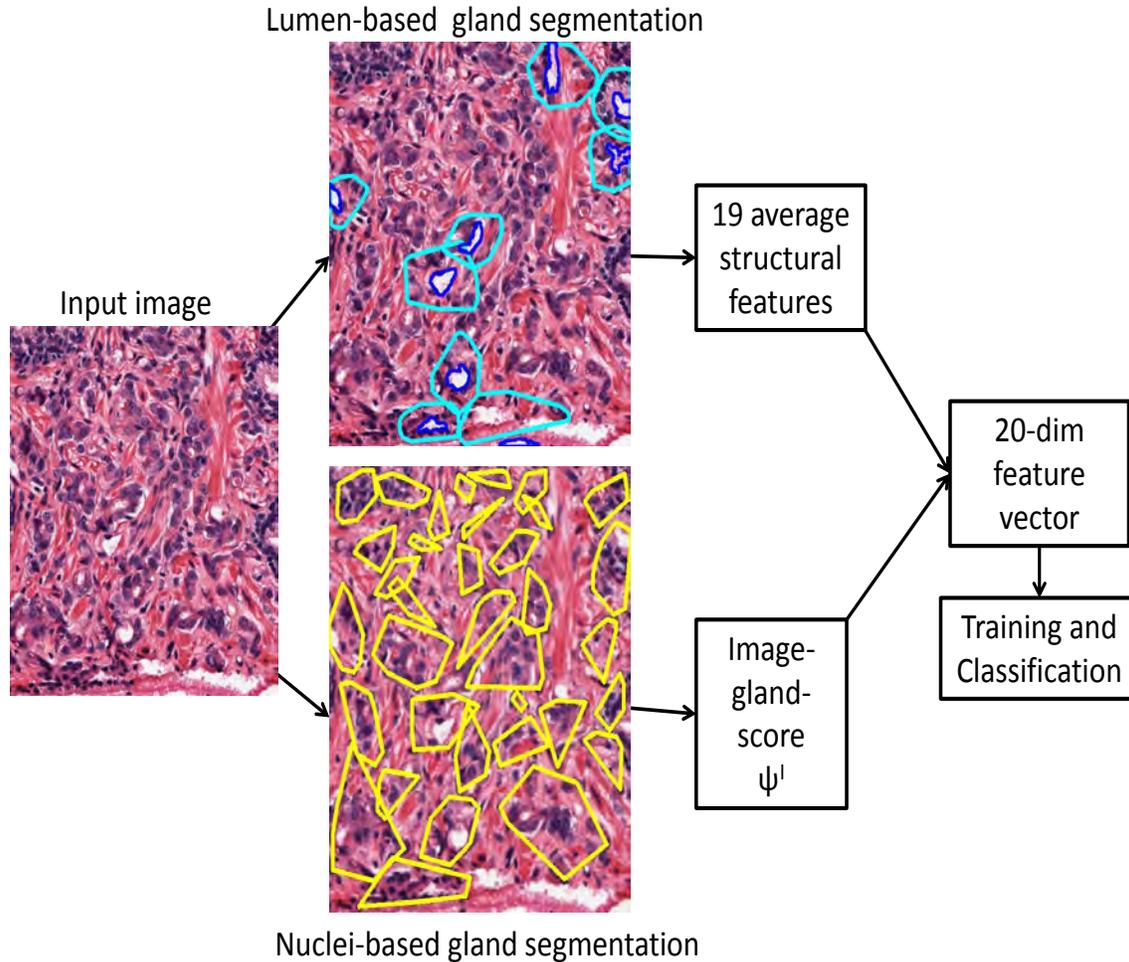
Figure 3.30: The fusion method for grade 3 vs grade 4 tissue image classification.

of the glands as the 19 image features, denoted as F_{19}[13].

Finally, we create a 20-dimensional feature vector $F_{20} = \{F_{19}, \psi^I\}$ to represent the image and use it for the grade 3 vs grade 4 tissue image classification problem. Figure 3.30 demonstrates the fusion method. We perform 10-fold cross validation (using SVM classifier with RBF kernel) on the grade 3 and grade 4 image database to evaluate this fusion method. The average classification accuracy obtained is 88.5% (with a standard deviation of 6.7%), which is an improvement over the result of the lumen-based method alone, 83.3% (7.1%) (Table 3.2). We further show the confusion matrices obtained by the lumen-based and the fusion

---

[13]We also identify and discard artifacts before computing the average features as done in section 3.3.
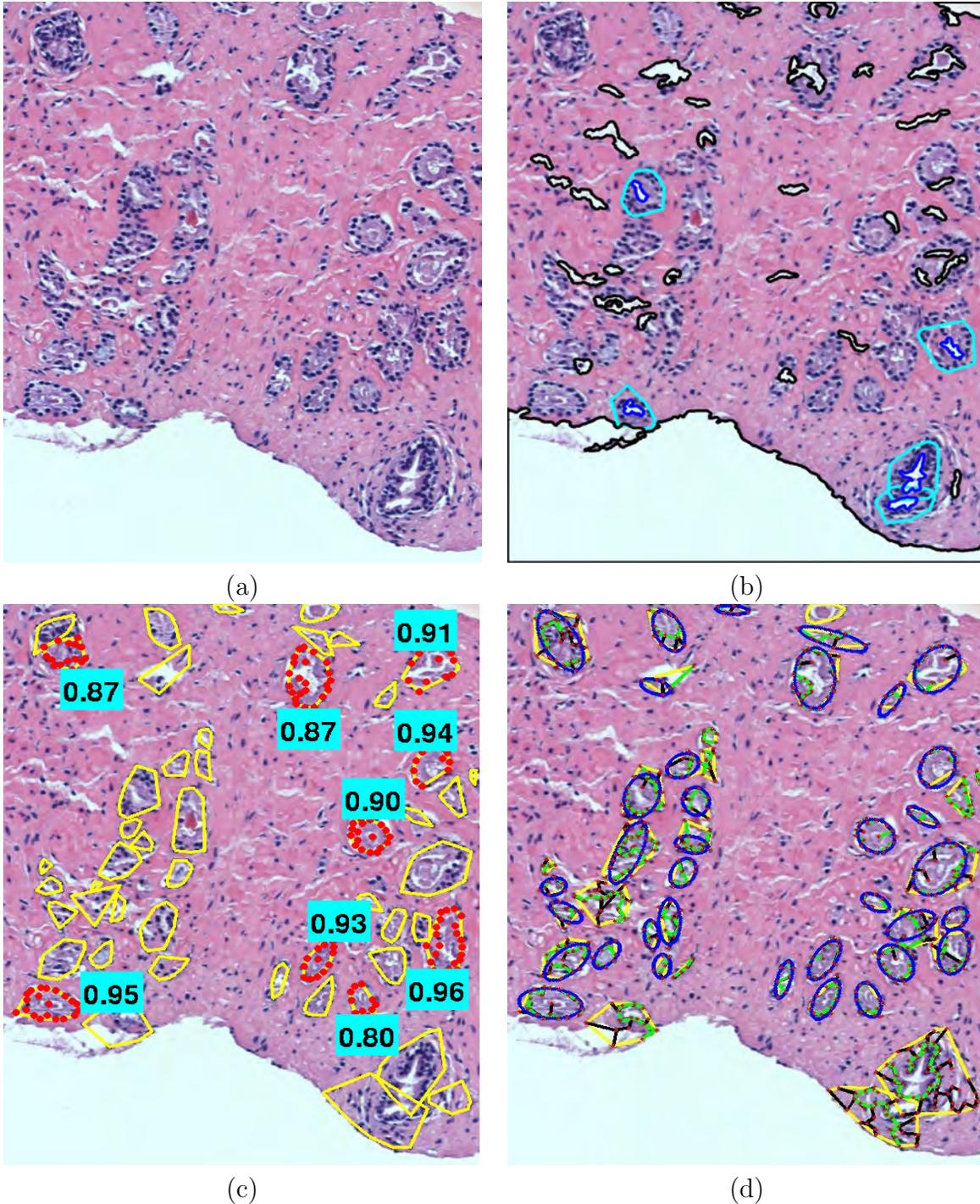
Figure 3.31: The usefulness of the nuclei-based method in grade 3 vs grade 4 tissue image classification. (a) A grade 3 image of $1,340 \times 1,030$ pixels (corresponding to a 0.67×0.51 mm^2 tissue region digitized at $20\times$). (b) The gland segmentation result of the lumen-based method (cyan contours), where detected lumina are shown as blue contours and detected artifacts are shown as black contours. (c) The gland segmentation result of the nuclei-based method, in which the segments with ψ values in the top 20% are indicated by red dots (nuclei in the segments), yielding $\psi^I = 0.9$. (d) The MSTs (green) and fitted ellipses (blue) computed for the segments in (c).

113

methods in Tables 3.6 and 3.7, respectively. The observations from these confusion matrices are the following:

1. For both the methods, the number of grade 4 images being misclassified is higher than the number of grade 3 images being misclassified. The reason is that in many grade 4 images, the cancer is still evolving from grade 3 to grade 4 patterns. Many glands in these images do not completely lose the gland structures, i.e., they still have a nuclei-forming closed chain structure on the boundary. Figure 3.32 shows examples of this problem. In the grade 4 images in this figure, although there are evidences of glands being fused with each other (indicated by green arrows), we can still see glands in which nuclei arrange as closed chain structures similar to glands in grade 3 images, instead of being randomly distributed (as in Figure 3.25b). Hence, these glands are segmented out by the nuclei-based method and obtain high ψ values (segments indicated by red dots). This means: (i) the nuclei-based gland segmentation method works properly and (ii) the misclassification is caused by the confusion in the gland structures, rather than by the segmentation results.

2. The main improvement of the fusion method over the lumen-based method is the identification of grade 3 images (there are only five grade 3 images being misclassified by the fusion method). This means that the nuclei-based method is able to efficiently detect the closed chain structure of the glands in grade 3 images, including the glands without detected lumen. An example showing this usefulness of the nuclei-based method is presented in Figure 3.31. In this figure, we show a grade 3 image that is misclassified as grade 4 when only the lumen-based method is used. The reason is that glands are not well-detected by the lumen-based method since only a few lumina are detected (Figure 3.31b). Moreover, the lumina in this image are very small, which is a similar feature to grade 4 images. However, when using the fusion method, we obtain the correct classification result because the nuclei-based method is able to detect the glands

114

without lumen, and obtain high ψ^I value for this image (Figure 3.31c).

		True label		
		Grade 3	Grade 4	Total
Predicted label	Grade 3	121	21	142
	Grade 4	13	49	62
	Total	134	70	204

Table 3.6: The confusion matrix obtained by the lumen-based method.

		True label		
		Grade 3	Grade 4	Total
Predicted label	Grade 3	129	18	147
	Grade 4	5	52	57
	Total	134	70	204

Table 3.7: The confusion matrix obtained by the fusion method.

3.5.6.5 Feature Weight

To estimate the importance of the features in the 20-dimensional feature vector used by the fusion method, we train a linear SVM classifier on the database used in this section and examine the weights of the features. Among all the features, the weight of the image-gland-score is the highest, which is two times higher than the weight of the second highest weight feature. This result shows the importance of the image-gland-score feature in the grade 3 vs grade 4 tissue image classification problem.

3.5.6.6 Low Magnification Images

We further conduct the same experiments discussed in this section (section 3.5.6) on $5\times$ magnification images, and obtain the distribution of ψ^I values for grade 3 and grade 4 images in Figure 3.33. This distribution corresponds to $\beta = 0.1$, at which the maximum mean difference $d_\mu^{5\times} = 0.16$ is obtained. So $d_\mu^{5\times} < d_\mu^{20\times}$ ($d_\mu^{20\times} = 0.24$ as mentioned earlier). Moreover, from Figures 3.27 and 3.33, we also observe a slightly larger discrimination between
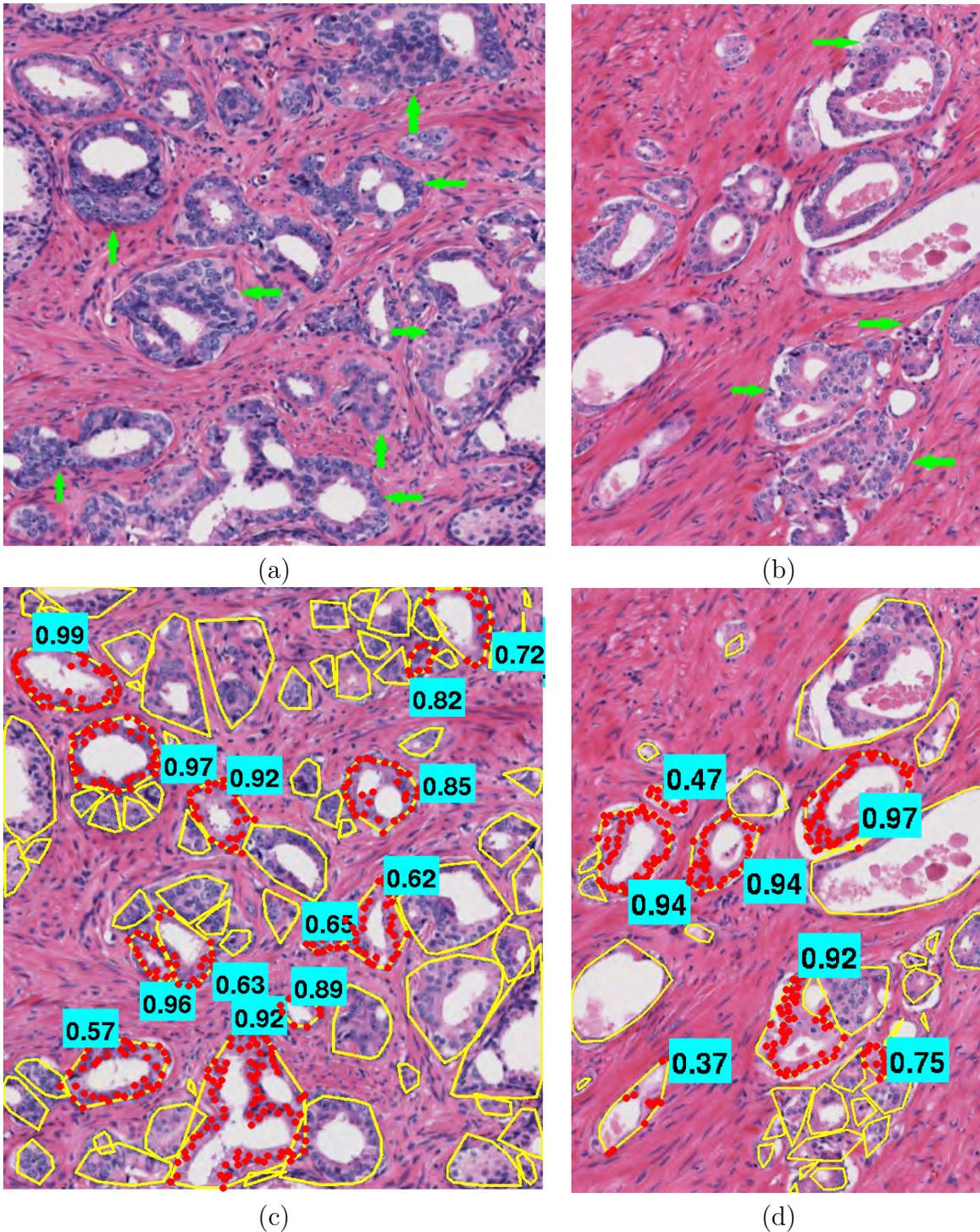
Figure 3.32: Examples of grade 4 images being misclassified as grade 3. (a) A grade 4 image of $1,500 \times 1,260$ pixels (corresponding to a 0.75×0.63 mm^2 tissue region digitized at 20×). (b) A grade 4 image of $1,760 \times 1,200$ pixels (corresponding to a 0.88×0.6 mm^2 tissue region digitized at 20×). (c) and (d) Gland segmentation results of the images in (a) and (b), respectively, in which the segments with ψ values in the top 20% are indicated by red dots (nuclei in the segments). The ψ values of these segments are shown in cyan boxes, yielding $\psi^I = 0.8$ for the image in (a) and $\psi^I = 0.77$ for the image in (b).

116
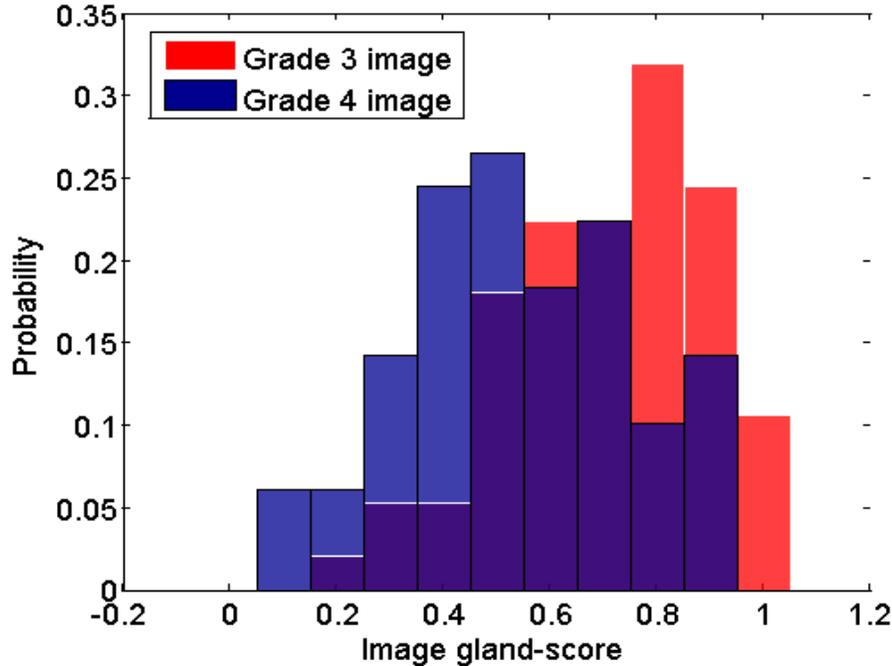
Figure 3.33: The distribution of ψ^I values computed for grade 3 and grade 4 images at $5\times$ magnification (obtained using $\beta = 0.1$).

the ψ^I values of grade 3 and grade 4 images at $20\times$ than at $5\times$. Finally, we also perform the grade 3 vs grade 4 tissue image classification problem using the fusion method on $5\times$ images, and obtain an accuracy of 86.3%, which is slightly lower than that for $20\times$ images (which is 88.5% as reported earlier).

The reason for the decreased performance of the nuclei-based method on $5\times$ images compared to $20\times$ images can be explained as follows. Recall that the nuclei-based method relies mostly on the detection of nuclei (using the radial-symmetry-based method), and the detection of lumen and stroma (using k-means algorithm). We observe that the detection results of nuclei and lumen at $5\times$ and $20\times$ images are comparable, since nuclei and lumen have salient color (this has been mentioned in section 3.4, Figure 3.3). However, the detection of stroma is more difficult because stroma color can be easily confused with cytoplasm color (this was previously mentioned in Figure 2.3). As a result, at low magnification images, which contain less information than high magnification images, the stroma detection results degrade. Moreover, the nuclei-based gland segmentation method is sensitive to the stroma

117

detection result because stroma directly affects the creation of links between the nuclei, which determine the nuclei-lumina-graph. In Figure 3.34, we show an example comparing the nuclei-based gland segmentation results for an image at 5× and 20× magnification[14]. Due to the misdetection of stroma at 5× compared to 20× (cyan circles in Figures 3.34b and 3.34e), some bad links that connect two different glands are created at 5×, but are not created at 20× (cyan circles in Figures 3.34c and 3.34d). As a result, these two glands are grouped together in the segmentation result at 5× but are separated in the segmentation result at 20× (cyan circles in Figures 3.34d and 3.34g).

3.6   Summary and Contributions

In this chapter, we have addressed the tissue image classification problem. As the first attempt, we applied a lumen-based method: performing gland segmentation using the NLA method (a lumen-based gland segmentation method) and using the average of the structural-contextual features extracted from the gland segments to perform the classification. The classification results obtained are higher than those of the popular texture-based methods used in the literature, for both low magnification (5×) and high magnification (20×) images. To further improve the grade 3 vs grade 4 classification result, which is a difficult problem, we have proposed the nuclei-based gland segmentation method and computed an image-gland-score to represent the image. The image-gland-score is combined with the features obtained from the lumen-based method to improve the result of the grade 3 vs grade 4 classification.

Unlike state-of-the-art methods (lumen-based methods), the nuclei-based gland segmentation method looks at the gland segmentation problem from a different perspective: it considers the glands as the groups of nuclei (or groups of nuclei and lumina), partially separated by stroma. As a result, it is able to detect glands without lumen, or glands with multiple lumina, which are the limitations of the lumen-based methods. In the nuclei-based gland segmentation method, we are able to utilize both the local information at the nuclei

---

[14]The 5× image is created by downsampling the original 20× image by four times
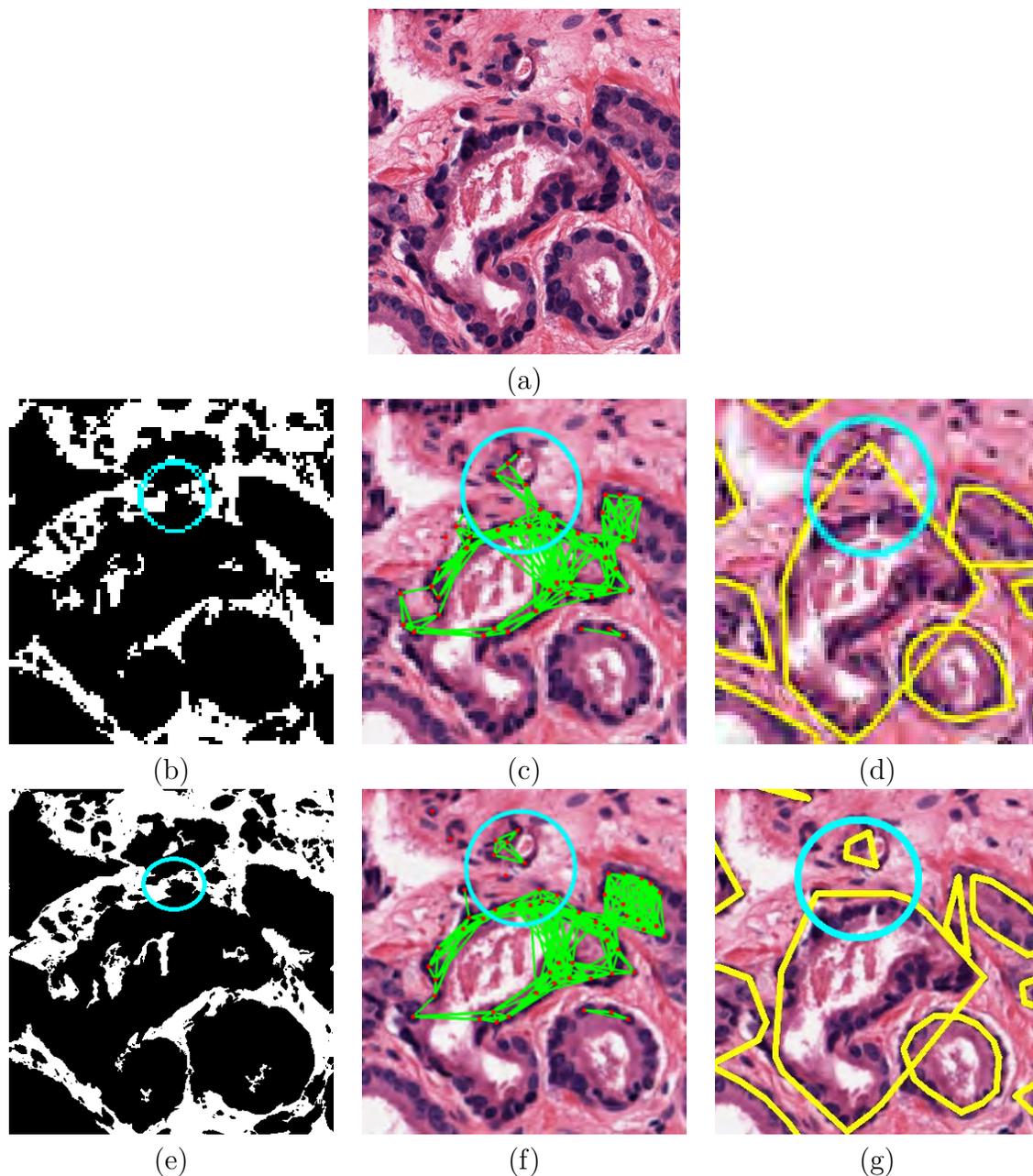
Figure 3.34: Comparison of nuclei-based gland segmentation results for an image at 5×
and 20× magnification. (a) A tissue image of 360 × 370 pixels, corresponding to a 0.18 ×
0.18 mm^2 tissue region digitized at 20×. Stroma detection result (white regions) at 5× (b)
and 20× (e). Links created (green lines) for the image at 5× (c) and 20× (f).
Segmentation result for the image at 5× (d) and 20× (g). The cyan circles in (b) and (e)
denote the area where stroma is not detected at 5× but being detected at 20×. This leads
to (c) bad links being created and (d) two different glands being grouped into one segment.
Such problems are not present at 20× ((f) and (g)).

119

and lumina (to construct the nuclei-lumina-graph), and the global information in the image (to find the weakest set of links in the entire graph using the normalized cut method). Moreover, we have proposed the closed chain structure measures and ellipse measures, which are used to estimate how similar the arrangement of a set of points is to the closed chain structure and the ellipse. These measures are combined in a learning framework to create a single gland-score, which describes how similar the arrangement of a nuclei-group is to the arrangement of nuclei in a gland. Next, an image-gland-score is derived and is shown to be a useful feature for the grade 3 vs grade 4 tissue image classification problem.

Finally, both the lumen-based and nuclei-based methods have been tested on both low magnification ($5\times$) and high magnification ($20\times$) images to evaluate the effect of image magnification on the performance of the methods.

Chapter 4

CONTENT-BASED PROSTATE TISSUE REGION RETRIEVAL

4.1 Introduction

4.1.1 Content-based Image Retrieval

Content-based image retrieval (CBIR) systems refer to the systems that are able to match a query image with images in a database by computing a similarity measure between the features (color, texture, shape, or a combination of them) of the images [79]. Due to the rapid development of digital cameras, storage devices, network capacity, etc, the amount of images being produced substantially increases, which makes it cumbersome to manually provide text description for all the images. Hence, CBIR systems become very useful, with a wide range of applications: a regular user searches for images of interest in the internet, the police search for images of criminals from a face image database or a fingerprint image database, a doctor searches for MRI images or CT images of similar cases, etc. Some of the first CBIR systems being developed include: the Kato system [80] supporting color and shape features, the IBM's QBIC system [81] supporting color, shape and texture features, and the Virage system [82] supporting color, color layout, texture and shape features.

The major challenges in developing a CBIR system are: (i) the high dimensionality in the feature space [83], (ii) the problem of automatic indexing of the images in the database [84], and (iii) the semantic gap between the low level features extracted from the images and the high level human interpretation of the images. For example, two images with similar color and texture can appear as dissimilar to the users. Inversely, two images with a large difference in the feature space can be interpreted as similar images by the users. Although developing a general CBIR system is very difficult (due to the semantic gap mentioned above), successful applications are available in special domains, e.g., Tattoo-ID [85] for tattoo

image recognition, VeggieVision system [86] for vegetable image recognition. In the medical domain, numerous CBIR studies have been presented to assist doctors and physicians in the diagnosis process.

4.1.2   CBIR in Medical Domain

A summary of the studies on medical image retrieval is given in Table 4.1. A more detailed survey on this topic can be found in [87]. While most of the medical image retrieval studies have performed query and retrieval on independent images, there is a large variation in the image databases that they used, i.e. the databases contain images taken from various types of imaging modalities (CT, MRI, tissue scanner, etc.), and taken from different organs in the body (brain, lung, breast, skin, etc.). Examples of these databases include: 15 positron emission tomography (PET) images at 128×128 pixels in [88], 57 breast cancer biopsy images at 640×480 pixel in [89], 196 axial brain images at 256×256 pixels in [90], 312 CT lung images (image sizes were not specified) in [69], 1,502 skin histology images at 1,280 $\times$ 1,024 pixels in [91], 11,000 radiographic images of different regions in the body (image size was not specified) in [70], 782 ultrasound images taken from abdominal organs (average image size is 640×480 pixels) in [68]. To evaluate the image retrieval system, images in the database were typically assigned to different categories. A retrieved image was considered to be relevant to the query if it belonged to the same category as the query image. For example, in [68], the authors divided the abdominal ultrasound images into 28 categories based on visually similar characteristics. Since the database was very large and heterogeneous in [70], the authors defined 116 categories corresponding to different body orientations, imaging modalities, body regions, and biological systems that the images belong to. In [91], the authors defined 18 categories corresponding to 18 biological structures appearing in their histology images.

The similarity between two images can be computed using the Euclidean distance between the corresponding feature vectors. Besides Euclidean distance, other distance measures such

as L_1 distance, Mahalanobis distance, city block distance, and histogram intersection have also been utilized (see [87] and Table 4.1).

4.1.3   Content-based Region Retrieval in Prostate Tissue Images

When a medical laboratory technician or a medical student who is not very experienced in Gleason grading wants to grade a region of interest (ROI) in a tissue slide image, it will be useful if automatic tools are available to assist him in this task. In chapter 3, we introduced an automatic tissue image classification module, which can automatically classify a ROI into different Gleason grades (normal, Gleason grade 3 and Gleason grade 4). In this chapter, we introduce a method that aims at searching for tissue regions that are similar to the ROI in a database of tissue slide images. Given this capability, the student (or technician) can search for tissue regions similar to the ROI that were annotated by experienced pathologists, and then use the Gleason grades of these tissue regions as references to grade the ROI. This is indeed feasible because regions that belong to the same Gleason pattern usually look similar (in terms of tissue texture, gland structures, etc.) to each other (Figure 4.1).

The objective of this chapter is to address the tissue region retrieval problem, which is illustrated in Figure 4.4. In this problem, given a query region, we need to search for tissue regions similar to the query region in a database of tissue slide images. Examples of a tissue slide image and a query region are delineated in Figures 4.2 and 4.3. We scan a window of the same size as the query region across the tissue slide images in the database and compare the image content of the window (test region) with the query region. Based on the similarity, we can rank the test regions from the most similar to the least similar and return the n most similar test regions to the user. For computational efficiency, we do not shift the window pixel by pixel in the database images but instead we choose a step size $\delta = 100$ pixels to shift the window in both row and column. This step size corresponds to the diameter of an average-sized gland in the tissue. This way we do not lose a significant amount of information while shifting the window.

Studies	Features used	Distance measure	Database	Accuracy
Beretti et al. [71]	Features extracted from the gray level co-occurrence matrix (GLCM)	Squared Euclidean distance	300 images representing a cellular body	An average precision of 0.77 and an average recall of 0.81
Caicedo et al. [91]	Color and gray level histograms, local binary patterns, bag of SIFT features, Sobel histogram	Tanimoto distance [92]	1,502 skin histology images	A precision of 0.68 at rank 1 retrieval
Huang et al. [93]	Color layout descriptor, Tamura features [94], edge histogram descriptor	Euclidean distance and L_1 distance	9,000 radio graphs	A precision of 0.81 at a recall of 0.1
Kwak et al. [68]	Gray level histogram, histogram moment, features extracted from the GLCM, wavelet transform	L_1 distance	782 ultrasound images taken from abdominal organs	A precision of 0.48 at the top 10% retrieved images
Rahman et al. [70]	Edge histogram descriptor, color layout descriptor, average gray value, features extracted from the GLCM	Cosine distance	11,000 radiographic images of different regions in the body	A precision of 0.75 at a recall of 0.1
Song et al. [95]	Bag of words model using Gabor filters	Weighted L_1 distance	1,134 PET-CT slice pairs	A precision of 0.83 at rank 3 retrieval
Sparks et al. [96]	Shape descriptor	Euclidean distance	91 DCE-MRI breast images	A value of 0.55 for the "area under the precision recall curve" measure
Wang et al. [97]	Wavelet transform	IRM metrics [98]	70,000 pathology image fragments	A precision of 0.9 at rank 50 retrieval

Table 4.1: Summary of studies on medical image retrieval.
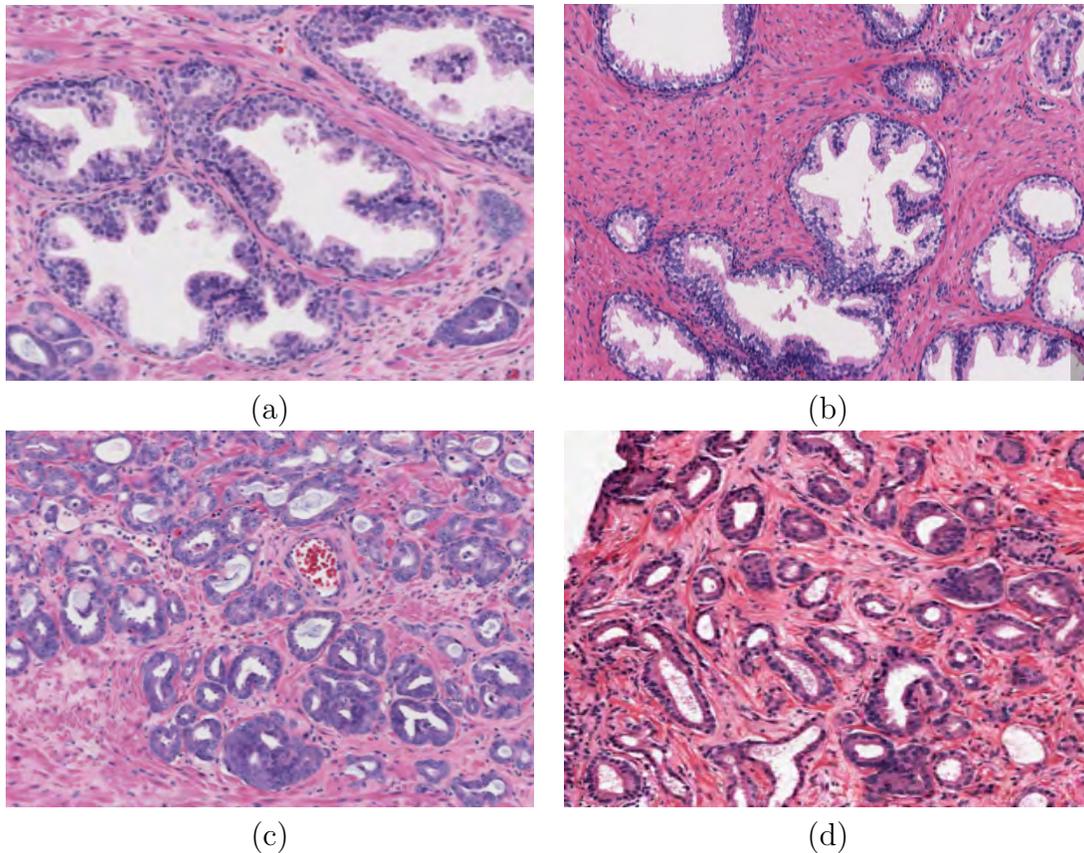
Figure 4.1: Correspondence between visual similarity of the tissue regions and their corresponding Gleason grades. Tissue regions that belong to normal patterns (a) and (b) are visually similar to each other (both contain large-sized glands). Tissue regions that belong to grade 3 patterns (c) and (d) are visually similar to each other (both contain small-sized glands). The average image size is 240×360 pixels, corresponding to a 0.5×0.72 mm^2 tissue region digitized at $5\times$ magnification.

In our region retrieval problem, we first need to compute the similarity between two tissue regions. To compute this similarity, we propose a *gland-based* method that utilizes glandular information in the region, including gland density, gland area, and gland structures.

4.2 Gland-based Region Similarity

The proposed method to compute the gland-based region similarity is illustrated in Figure 4.5. To implement this region similarity method, we use the NLA method discussed in chapter 2 to segment glands from the image and compute the gland feature vector $\mathbf{f}$, including

125
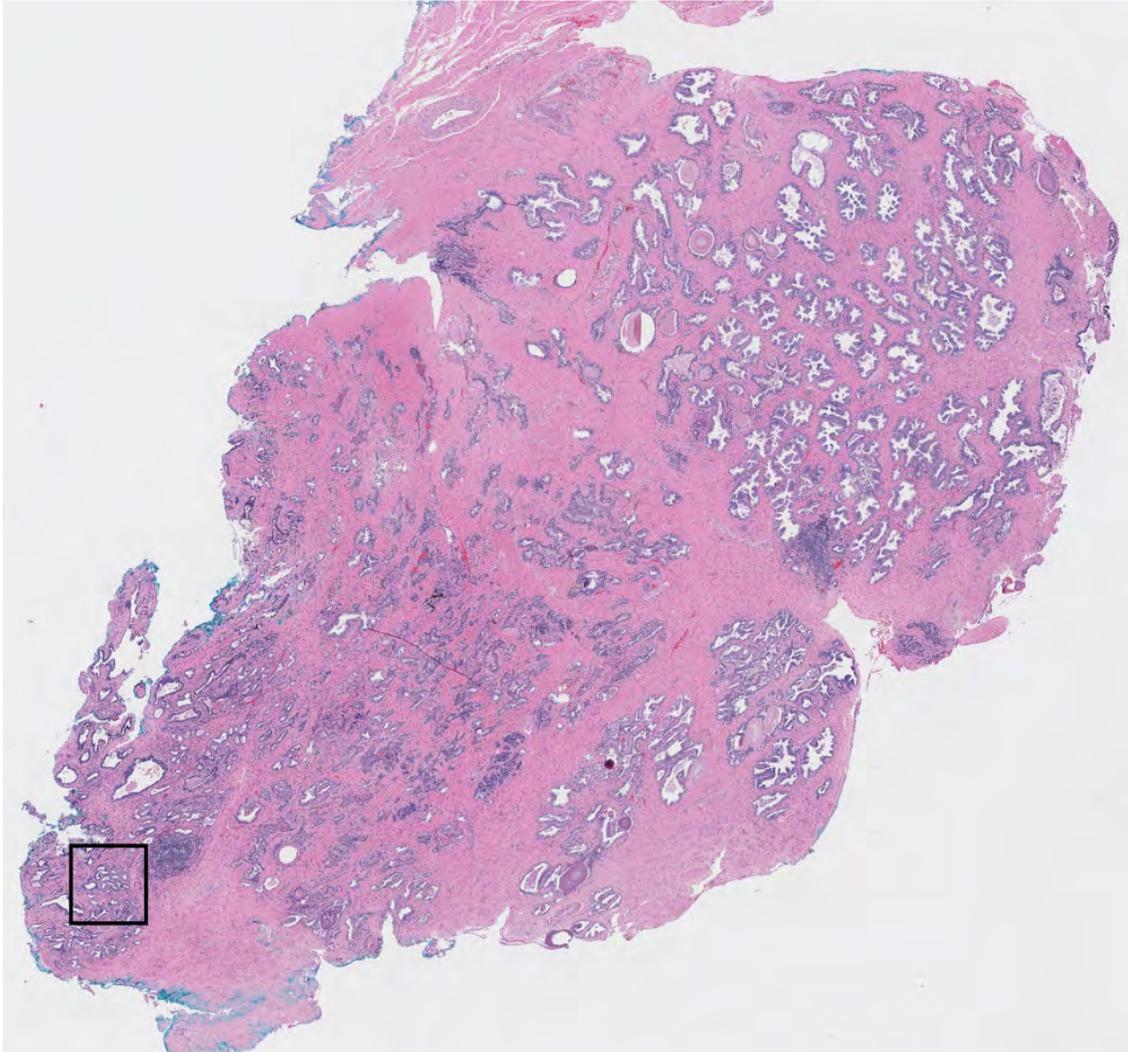
Figure 4.2: A tissue slide image and a subimage identified as a query region (black rectangle). The image size is $5{,}000 \times 4{,}500$ pixels, corresponding to a 10×9 mm^2 tissue slide digitized at $5\times$ magnification.
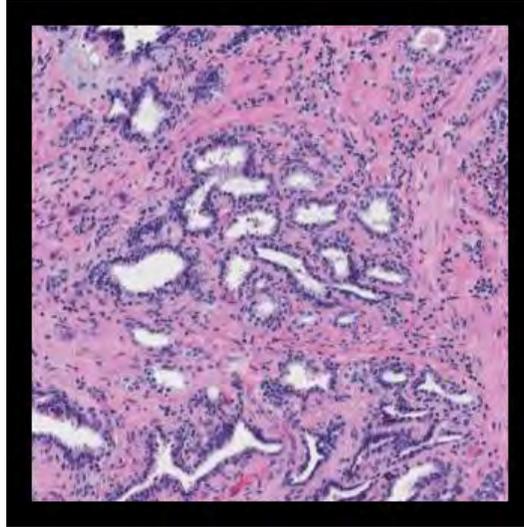
Figure 4.3: Close-up of the query region in Figure 4.2. The size of the region is 350×350 pixels, corresponding to a 0.7×0.7 mm^2 tissue region digitized at $5\times$ magnification.

Find regions similar to the query region
in a database of annotated tissue slide images

A window is scanned across the images to find regions similar to the query

Annotation

New tissue slide image

Compute similarity

Query region

Tissue Slide Image A   Tissue Slide Image B   Tissue Slide Image C

Retrieval results

Ranked list of similar regions based on their similarity to the query region (high to low)

1   2   3   .....
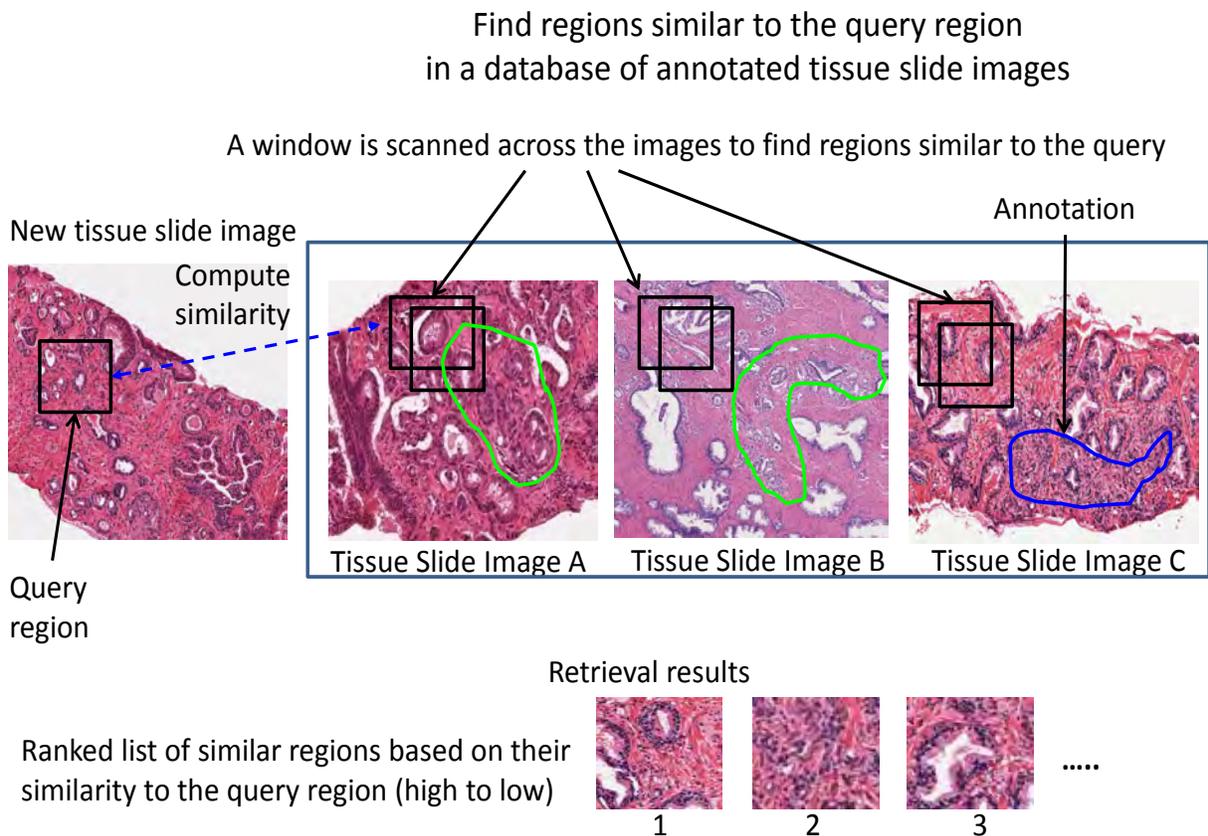
Figure 4.4: The tissue region retrieval problem. A user can search for regions similar to a query region in the tissue slide image database that has been annotated with Gleason grades. The Gleason grades of the retrieved regions can serve as the reference to grade the query region.

22 structural and contextual features (described in Table 2.2). We do not use the nuclei-based gland segmentation method (presented in chapter 3) for this task because:

1. When applying the nuclei-based gland segmentation method on grade 4 tissue regions, the segments we obtain are more likely to contain randomly distributed nuclei and random structures than contain individual gland units with well-defined structures. Hence, we are unable to compute the similarity between these segments.

2. The nuclei-based gland segmentation method is computationally more expensive than the NLA method since it requires several steps: nuclei detection, graph construction, recursive normalized cuts and gland-score computation. This method was shown to be useful in the tissue image classification problem since we only processed small tissue images (average size is 350×345 pixels at $5\times$ magnification) in that problem. The tissue images used in this retrieval problem, in contrast, are very large (average image size is $2,800 \times 4,500$ pixels at $5\times$), which significantly increases the computation time of the method.

3. Using the NLA method and the gland feature vector $\mathbf{f}$ mentioned above, we obtained a very high accuracy for the normal vs cancer, and a relatively good accuracy for the grade 3 vs grade 4 tissue image classification problems (Table 3.2). Hence, this is the suitable solution for this region retrieval problem.

There are three types of glandular similarities that can be computed between two regions: (i) similarity with respect to gland density S_n, (ii) similarity with respect to gland area S_a, and (iii) similarity with respect to gland structures S_s. The computation of S_n, S_a and S_s is described in Figure 4.5. Additional details of computing S_s is provided in Algorithm 4.1. To perform Algorithm 4.1, we need to compute a similarity score for a pair of glands. This similarity score should be able to describe the similarity between the two glands with respect to their structural properties. As mentioned in chapter 2, the gland feature vector $\mathbf{f}$ (a 22-dimensional vector) describes all the structural information of the gland, including

information about the cytoplasm, nuclei, lumen, gland morphology, and also describes contextual information. Hence, we can compute the similarity score for the two glands (g, g') as:

$$s(g, g') = \frac{1}{\|\mathbf{f} - \mathbf{f}'\|_2}. \tag{4.1}$$

Since the ranges of different gland features are different, e.g. the nuclei density feature is in the range [0,1] while the gland area feature can be in a much larger range (e.g. [30, 240,000] pixels), we normalize all gland features so that they all have zero mean and unit variance before computing the Euclidean distance. In Algorithm 4.1, we first compute the similarity scores for all pairs of glands between the query region Q and the test region T. Next, we sort the gland pairs based on the similarity scores. For each gland g_i in the query region Q, we associate a weight w_i to g_i. The weights of all glands in Q, $\mathbf{w} = \{w_1, w_2, \ldots, w_{n_Q}\}$, are computed based on their areas. Our purpose in using these weights is that a large gland should contribute more to the structural similarity between the two regions than a small gland. Moreover, we have more confidence in using larger area glands since they are less likely to be noisy regions. By letting $\{a_1, a_2, \ldots, a_{n_Q}\}$ denote the areas of all the glands in Q, we first compute the minimum area, $a_m = min\{a_i\}_{i=1}^{n_Q}$, followed by the weights:

$$w_i = log\left(\frac{a_i}{a_m}\right) + 1, \forall i \in [1, n_Q]. \tag{4.2}$$

The weights satisfy the following properties:

- The weight of the smallest gland is 1 and the weights of all the other glands are greater than 1, $w_m = 1$, and $w_i > 1$, $\forall i \neq m$.

- If $a_i > a_j$, we also have $w_i > w_j$.

The log transformation is used because we do not want a linear relationship between the weights and the gland areas. More precisely, we do not want the differences in weights to be as large as the differences in gland areas (note that gland area has a large variance). The final similarity score is computed as $S_g = S_n \times S_a \times S_s$.

Goal: Compute the region similarity S_g between a query region and a test region using glandular information

Query region Q

Test region T

score₁

score₂

score₃

score₄

$\mathbf{q} = \{q_i\}$, $\mathbf{t} = \{t_j\}$
$(q_i, t_j$: glands in Q, T)
$k = 1$
While ($\mathbf{q} \neq \emptyset$ and $\mathbf{t} \neq \emptyset$)
  1. $(q_m, t_n) = \text{argmax}_{i,j}\, s(q_i, t_j)$
     $(\forall q_i \in \mathbf{q},\, t_j \in \mathbf{t})$
     $(s(q_i, t_j)$: See Equation 3.1)
  2. $\text{score}_k = w_m s(q_m, t_n)$
     $(w_m$: See Equation 3.2)
  3. $k = k + 1$
  4. $\mathbf{q} = \mathbf{q} \setminus \{q_m\}$, $\mathbf{t} = \mathbf{t} \setminus \{t_n\}$
end

$$S_n = \frac{\min(n_Q, n_T)}{\max(n_Q, n_T)}$$

Region similarity w.r.t. gland density
(n_Q, n_T: number of glands in Q, T)

$$S_a = \frac{\min(a_Q, a_T)}{\max(a_Q, a_T)}$$

Region similarity w.r.t. gland area
(a_Q, a_T: total gland area in Q, T)

$$S_s = \text{mean}(\text{score}_k)$$

Region similarity w.r.t. gland structures

$$S_g = S_n \times S_a \times S_s$$

Figure 4.5: The gland-based method for computing the similarity S_g between a query region Q and a test region T.

4.3 Database of Tissue Slide Images

As mentioned in chapter 1, the database used in this problem includes 39 tissue slide images at $5\times$ magnification (average image size is $2,800 \times 4,500$ pixels). In each tissue slide image, we manually circumscribe different areas, each of which corresponds to one of the three Gleason grades (normal, grade 3, and grade 4). This markup is done based on the annotations provided by the pathologist. An example of a tissue slide image with this markup is shown in Figure 4.6. Based on these circumscribed areas, we can assign labels to both query regions

---

Algorithm 4.1 Computing region similarity based on structural information

---

Input: Query region $Q = \{q_1, q_2, \ldots, q_{n_Q}\}$ and test region $T = \{t_1, t_2, \ldots, t_{n_T}\}$ (q_i, t_i denote the glands in the query and test regions, respectively). Weights of glands in Q, $\mathbf{w} = \{w_1, w_2, \ldots, w_{n_Q}\}$. Gland similarity scores $s(q_i, t_j) \ \forall q_i \in Q, t_j \in T$ ($s(q_i, t_j)$ is computed by Equation 4.1)

Output: Similarity between Q and T with respect to gland structures, denoted by S_s

  1: $n^* \leftarrow \min(n_Q, n_T)$
  2: **for** $k = 1$ to n^* **do**
  3:   Select the pair (q_m, t_n), where $q_m \in Q$ and $t_n \in T$ such that $s(q_m, t_n) > s(q_i, t_j)$ $\forall q_i \in Q, t_j \in T$ and $q_i \neq q_m, t_j \neq t_n$. We consider q_m and t_n as the matching glands at this iteration
  4:   $s_k \leftarrow w_m s(q_m, t_n)$
  5:   $Q \leftarrow Q \backslash q_m, T \leftarrow T \backslash t_n$ (q_m and t_n will not be considered again)
  6: **end for**
  7: $S_s \leftarrow \mathrm{mean}(s_k) \ \forall k \in [1, n]$

---

and test regions as follows. Let p_1, p_2 and p_3 denote the percentages of the region that contain normal, grade 3, and grade 4 areas, respectively. We assign a label m to the region where $m = \mathrm{argmax}_i p_i$, $i \in [1, 2, 3]$ and $p_m \geq 0.25$. On the other hand, if $p_i < 0.25, \forall i \in [1, 2, 3]$, we assign a label 0 (background) to the region. This means that we assign a Gleason grade to the region only if more than 25% of the region area contains the corresponding Gleason grade. In general, a region can be assigned a label in the set {0, 1, 2, 3}, which correspond to background, normal, grade 3, and grade 4 regions, respectively. In Figure 4.6, we also give two examples of region labeling: the region inside the black rectangle is labeled as background and the region inside the red rectangle is labeled as grade 3.

Similar to the tissue image classification problem, in this retrieval problem, we also need to perform the preprocessing step to remove artifacts from the tissue slide images. In order to achieve this, we select nine images (out of 39 images in the database) that contain artifacts, normal and cancer glands and contain large variations in color information to use as training images. From these training images, we select 190 artifacts and 220 glands (100 normal and 120 cancer glands), and learn an SVM classifier to separate artifact vs gland. The remaining 30 images are used for query and retrieval, which is denoted as the query-retrieval set. The trained SVM classifier that separates artifacts vs glands is applied on the
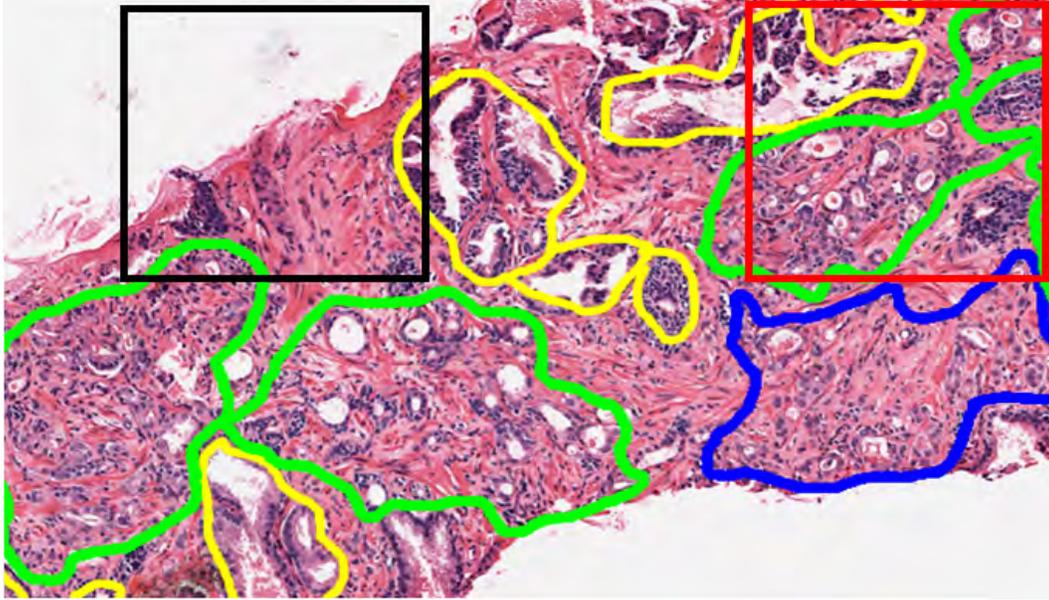
Figure 4.6: A part of a tissue slide image with circumscribed areas of different Gleason grades. Yellow denotes normal areas, green denotes grade 3 areas, and blue denotes grade 4 areas. The black rectangle on the left denotes a region that is assigned a label 0 (background) because only a small part of the region contains normal and grade 3 areas. On the other hand, the red rectangle on the right denotes a region that is assigned a label 2 (grade 3) because a substantial part of the region (more than 25%) contains grade 3 area. The image size is 420×680 pixels, corresponding to a 0.84×1.36 mm^2 tissue region digitized at $5\times$ magnification.

images in the query-retrieval set to remove artifacts from these images before performing the gland-based region retrieval method. The selection of training images does not affect the performance of the proposed method as long as these training images contain normal, cancer glands and artifacts (we can find artifacts in almost any image), and contain large variations in color. This is because the artifact vs gland classification is an easy problem (we obtained a high accuracy for this problem as mentioned in chapter 2).

4.4 Region Retrieval Methods

We compare the performance of six different region retrieval methods: texture-based, bag of words (BoW) based, category-based, gland-based, and two fusion methods. The first fusion method is the combination of the BoW-based and the gland-based methods (which is referred

Method	Strategy to compute the similarity between a query region Q and a test region T
Texture-based	$S_t = \dfrac{1}{\|\mathbf{F}_Q - \mathbf{F}_T\|_2}$, where F_Q and F_T denote the normalized textural feature vectors of Q and T (the features are normalized so that they all have zero mean and unit variance)
BoW-based	$S_b = \sum_k min(h_k^Q, h_k^T)$, where h_k^Q and h_k^T denote the frequencies at the kth bin of the texton histograms of Q and T, respectively
Category-based	$S_c = \dfrac{1}{D_{KL}(\mathbf{P}_Q \| \mathbf{P}_T)}$, where P_Q and P_T denote the category confidence vectors of Q and T; D_{KL} denotes the KL divergence
Gland-based	The computation of the gland-based similarity S_g is described in section 4.2
BoW-gland fusion	$S_{bg} = S_b \times S_g$
Category-gland fusion	$S_{cg} = S_c \times S_g$

Table 4.2: The six methods to compute region similarity, which we use for comparison.

to as the BoW-gland fusion method), while the second fusion method is the combination of the category-based and the gland-based methods (which is referred to as the category-gland fusion method).

In the texture-based method, to compute the similarity between two regions, we use the reciprocal of the Euclidean distance between the textural feature vectors extracted from the two regions. The textural features that we use include three types of features: first order statistics of pixel intensity, features extracted from the GLCM [4] (which was used in [68–71]), and Gabor filter features [99] (which was used in [100–103]), yielding a 111-dimensional feature vector.

In the category-based method (which was used in [91] and [70]), we first train an SVM classifier to separate four different categories of regions (background, normal, grade 3, and grade 4). This is done using an independent set of training regions, which include 1,397 background, 1,778 normal, 164 grade 3, and 191 grade 4 regions (these regions are extracted from the nine training images described earlier). The trained SVM classifier is used to compute

Evaluation measure	Texture-based	BoW-based	Category-based	Gland-based	BoW-gland fusion	Category-gland fusion
AUPRC	0.39	0.43	0.44	0.47	0.48	**0.49**
$Relevant_{10}$	5.39	6.15	5.43	6.77	**6.87**	6.01

Table 4.3: The values of AUPRC and $Relevant_{10}$ measures for the six retrieval methods in Table 4.2.

the category confidence (probability) vector $\mathbf{P} = \{P(\omega_1|\mathbf{F}), P(\omega_2|\mathbf{F}), P(\omega_3|\mathbf{F}), P(\omega_4|\mathbf{F})\}$ of the query and test regions, where $\mathbf{F}$ denotes the textural feature vector (we use the same 111 textural features used in the texture-based method in this case) and $\omega_1, \omega_2, \omega_3$, and ω_4 denote the background, normal, grade 3, and grade 4 categories, respectively. The $\mathbf{P}$ vector describes the probability that the region belongs to each of the four categories. In this method, to compute the similarity between two regions, we use the Kullback-Leibler (KL) divergence, which is a common measure for probability-distance (used in [104], [105], and [106]).

For the BoW method, we use the procedure described in [40] (which has been mentioned in chapter 3, section 3.2) to implement the method. In this method, a *texton* histogram [40] is computed for each region. Hence, we estimate the similarity between two regions by computing the histogram intersection between two histograms of the two regions. Table 4.2 summarizes the six region retrieval methods.

4.5 Experimental Results

4.5.1 Evaluation Method

From the query-retrieval set (including 30 tissue slide images), we select 430 query regions (193 normal, 170 grade 3, and 67 grade 4 regions) in such a way that these regions cover most of the variations in the glandular structures in all the tissue slide images and do not overlap each other. The size of the query region and of all the test regions is 350×350
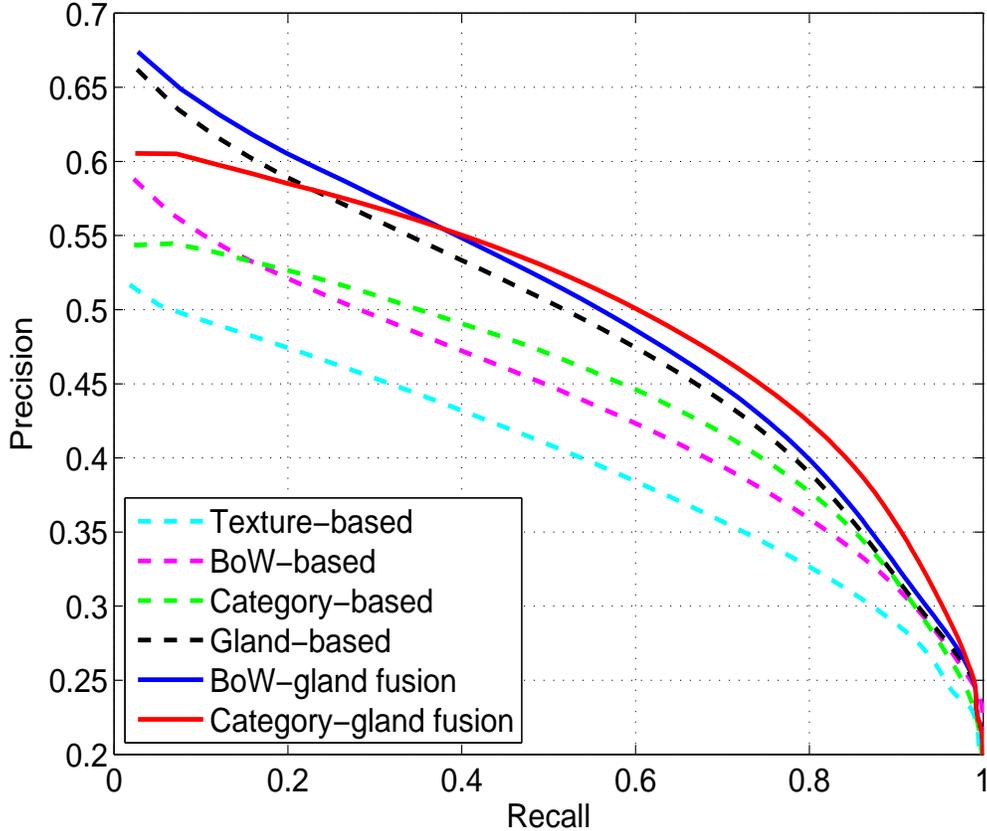
Figure 4.7: The precision-recall curves of the six region retrieval methods in Table 4.2.

pixels. This size is chosen because such a region can completely contain a very large gland with a diameter of approximately 300 pixels. For each of the six methods mentioned in Table 4.2, we use the following evaluation protocol.

For a query region Q that belongs to tissue slide image k, we compute the similarity scores between Q and the test regions in all the 29 tissue slide images other than the tissue slide image k. Let n_Q denote the total number of test regions in this case. The average value of n_Q for all the query regions is 33,630, which includes 17,767 background, 12,014 normal, 3,022 grade 3, and 826 grade 4 regions. Based on the region similarity scores computed, we sort the n_Q test regions from the highest similarity to the lowest similarity. Based on the resulting ranking, we compute the precision and recall for the top t regions among the n_Q regions as follows:

- Precision: $p_t = \dfrac{Relevant_t}{t}$, where $Relevant_t$ denotes the number of relevant regions

135

Evaluation measure	Texture-based	BoW-based	Category-based	Gland-based	BoW-gland fusion	Category-gland fusion
AUPRC	0.44	0.50	0.52	0.54	0.55	**0.57**
$Relevant_{10}$	6.75	7.53	6.86	7.88	**7.98**	7.37

Table 4.4: The values of AUPRC and $Relevant_{10}$ measures for the six region retrieval methods for normal vs cancer regions.

    among the top t retrieved regions (a test region is considered relevant if it has the same label as the query region). Intuitively, the precision indicates the retrieval accuracy.

- Recall: $r_t = \dfrac{Relevant_t}{Relevant_n}$, where $Relevant_n$ denotes the total number of relevant regions among all the n_Q test regions. Intuitively, the recall measures the percentage of all the relevant regions in the database that are retrieved.

By varying the value of t from 1 to n_Q, we obtain n_Q pairs of (recall, precision) values. The average precision and recall values obtained for all the query regions is plotted as a the precision-recall curve. The area under the precision-recall curve (AUPRC) is reported. A larger AUPRC value corresponds to a better retrieval performance. For a more intuitive measure of retrieval performance, we also report the average number of relevant regions among the top 10 retrieved regions (we refer to this measure as $Relevant_{10}$).

4.5.2 Overall Retrieval Results

The retrieval results of the six methods described in Table 4.2 are shown in Table 4.3 and Figure 4.7. These results show that the BoW-gland fusion method has the best accuracy for the top 10 retrieved regions, while the category-gland fusion method has the best performance for the precision vs recall measures.
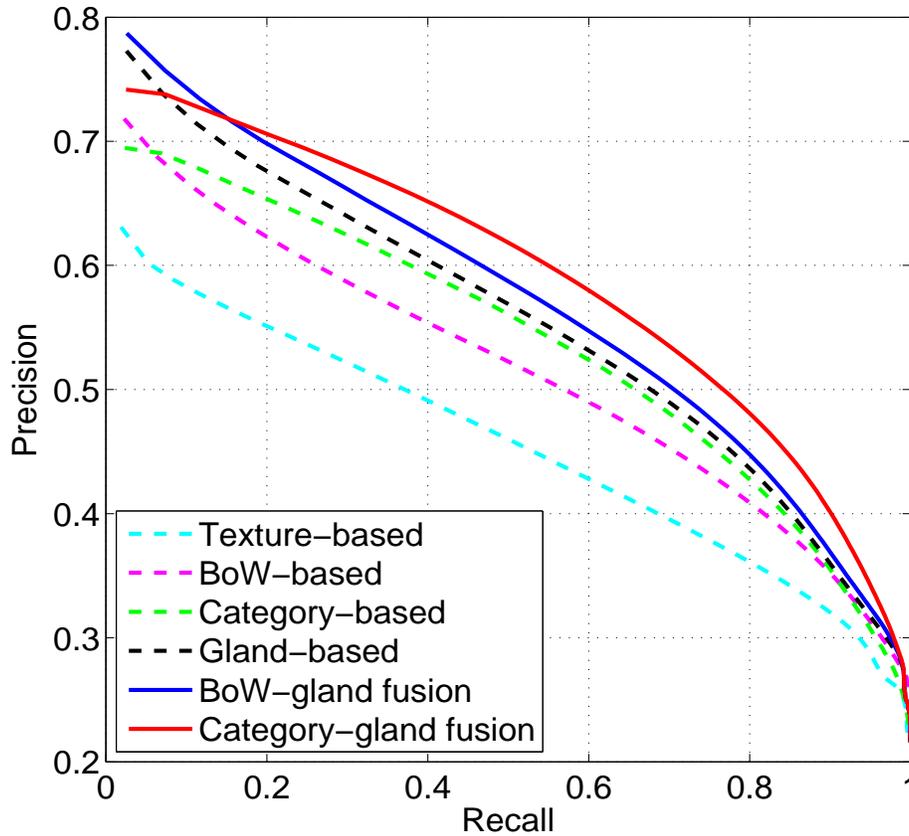
Figure 4.8: The precision-recall curves of the six region retrieval methods for normal vs cancer region retrieval.

4.5.3 Normal vs Cancer Retrieval Results

Due to the continuous growth of cancer in the tissue, it is often difficult to discriminate between grade 3 and grade 4 regions. Hence, in this experiment, we consider the normal vs cancer region retrieval results; where all grade 3 and grade 4 regions are labeled as cancer regions. The results of the six retrieval methods for normal vs cancer region retrieval are shown in Figure 4.8 and Table 4.4. Compared to the previous experiment (which considers grade 3 and grade 4 as different labels), the performance of all the methods improve. The BoW-gland fusion method still obtains the best result for the $Relevant_{10}$ measure, and the category-gland fusion method still obtains the best result for the precision vs recall measures.

137
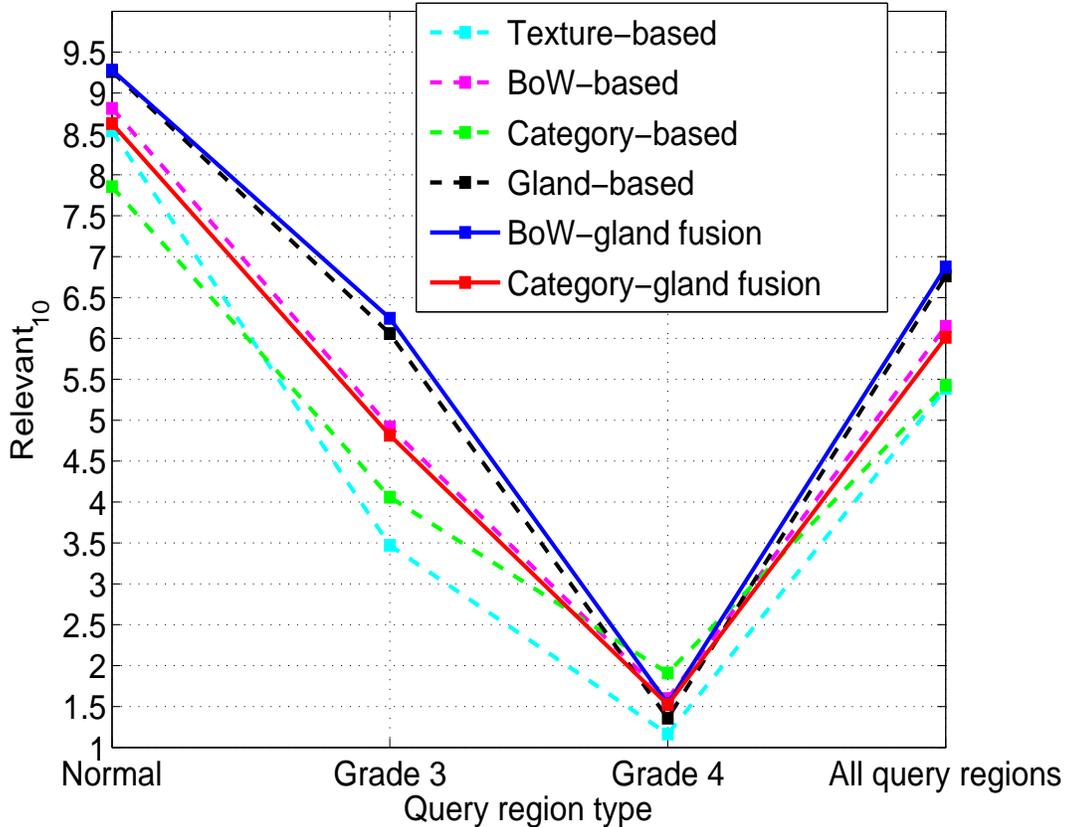
Figure 4.9: The values of the $Relevant_{10}$ measure obtained by the six region retrieval methods for specific query region types and for all query regions taken together.

4.5.4 Performance for Different Query Region Types

In this experiment, we analyze the performance of the six region retrieval methods for different types of query regions. As mentioned earlier, there are three types of query regions considered: normal, grade 3, and grade 4. Figure 4.9 shows the values of the $Relevant_{10}$ measure obtained by the six region retrieval methods for each query type, and for all query regions taken together. For all methods, the performance for the normal query region is the highest, followed by the performance for the grade 3 query region. The performance for the grade 4 query region is the worst. Furthermore, the BoW-gland fusion method performs the best for the normal and grade 3 query regions, while the category-based method performs the best for the grade 4 query regions.

4.5.5 Reject Option

As shown in Figure 4.9, the retrieval results for grade 4 regions are the lowest. The reason is that, in grade 4 regions, glands are poorly-defined, or even occluded, so the gland segmentation and gland feature extraction procedures become less reliable. Hence, we are interested in analyzing the retrieval results of the retrieval methods with respect to the total gland area detected in the tissue region. In Figure 4.10, we plot the $Relevant_{10}$ values of the gland-based method, and the two fusion methods (which obtain the best performance in the previous experiments) for query regions of different ranges of total gland area, e.g., the $Relevant_{10}$ value for the range [0, 3,000] pixels in this figure is the average $Relevant_{10}$ values for the query regions whose total gland area falls within this range. It is obvious from the figure that the larger the total gland area in the query region is, the more confidence we have with the retrieval results. Therefore, we can apply a rejection rule on the input query region, i.e., if the total gland area in the query region is less than a threshold R_t, we reject this region, otherwise, we perform the retrieval. In Figure 4.11, we plot the $Relevant_{10}$ values of the gland-based and the BoW-gland fusion methods (which obtain best results in Figure 4.10) with respect to different values of the rejection threshold R_t.
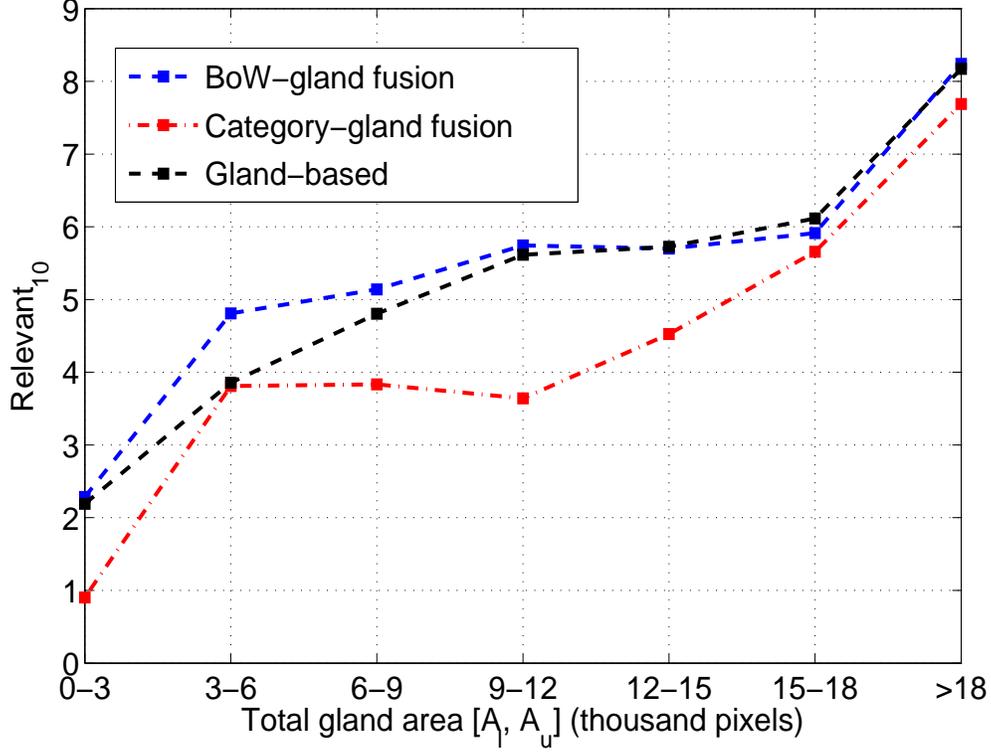
Figure 4.10: Plot of the $Relevant_{10}$ values obtained by the gland-based and the two fusion methods for different ranges $[A_l, A_u]$ of total gland area of the query regions.

4.5.6 Retrieval Examples

Retrieval examples of the gland-based method are shown in Figures 4.12, 4.13, and 4.14. In these examples, we show three different types of query regions and the top five non-overlapping retrieved regions (recall that the retrieved regions can overlap due to the shifting of the window in the tissue slide images) with highest similarity to the query region. The normalized similarity score (which is in the range [0,1]) is also shown for each of the five retrieved regions. We compute the normalized score as follows. Let $S_1, S_2, \ldots, S_5$ denote the original similarity scores between the query and the five retrieved regions computed by the gland-based method ($S_1 > S_2 > \cdots > S_5$); the normalized similarity scores $S_1^N, S_2^N, \ldots, S_5^N$ are computed as: (i) $S_1^N = 1$, (ii) $S_i^N = S_i^N/S_1^N$, $\forall i \in [2, 5]$. As shown in Figures 4.12, 4.13, and 4.14, in general, the normalized similarity scores obtained for the normal query region are the highest, followed by the scores obtained for the grade 3 query region, and the scores

140
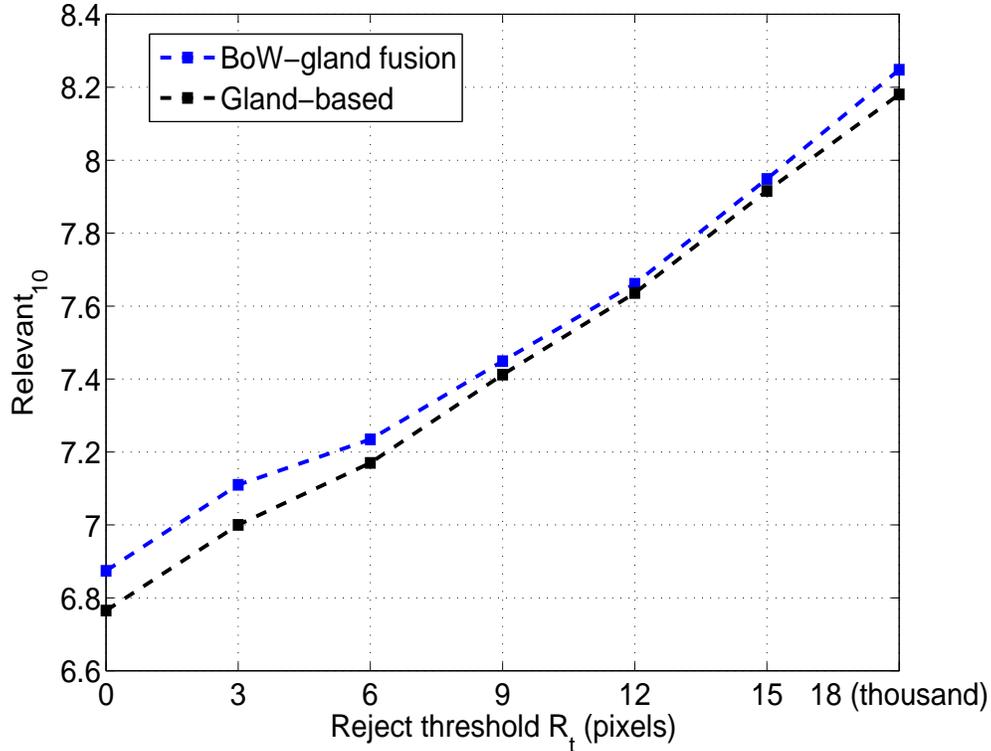
Figure 4.11: Plot of the $Relevant_{10}$ values obtained by the gland-based and the BoW-gland fusion methods for different values of the rejection threshold R_t.

for the grade 4 query region being the lowest. This means that we have the most confidence with the retrieval results of a normal query region and the least confidence with the retrieval results of a grade 4 query region.

For the normal query region in Figure 4.12, all five retrieved regions are relevant, i.e., labeled as normal. For the grade 3 query region in Figure 4.13, the rank 1 and rank 3 retrieved regions are normal regions, but they contain many small glands similar to the query region. Hence, they have high similarity scores to the query region. For the grade 4 query region in Figure 4.14, all the five retrieved regions have many small glands, which are similar to the query region. However, only the rank 1 retrieved region has glands that are fused together, while the remaining 4 retrieved regions do not have that property. As a result, only the rank 1 retrieved region is considered as a grade 4 region like the query region.

Query region	Rank 1 (1)	Rank 2 (0.99)
Rank 3 (0.98)	Rank 4 (0.97)	Rank 5 (0.97)
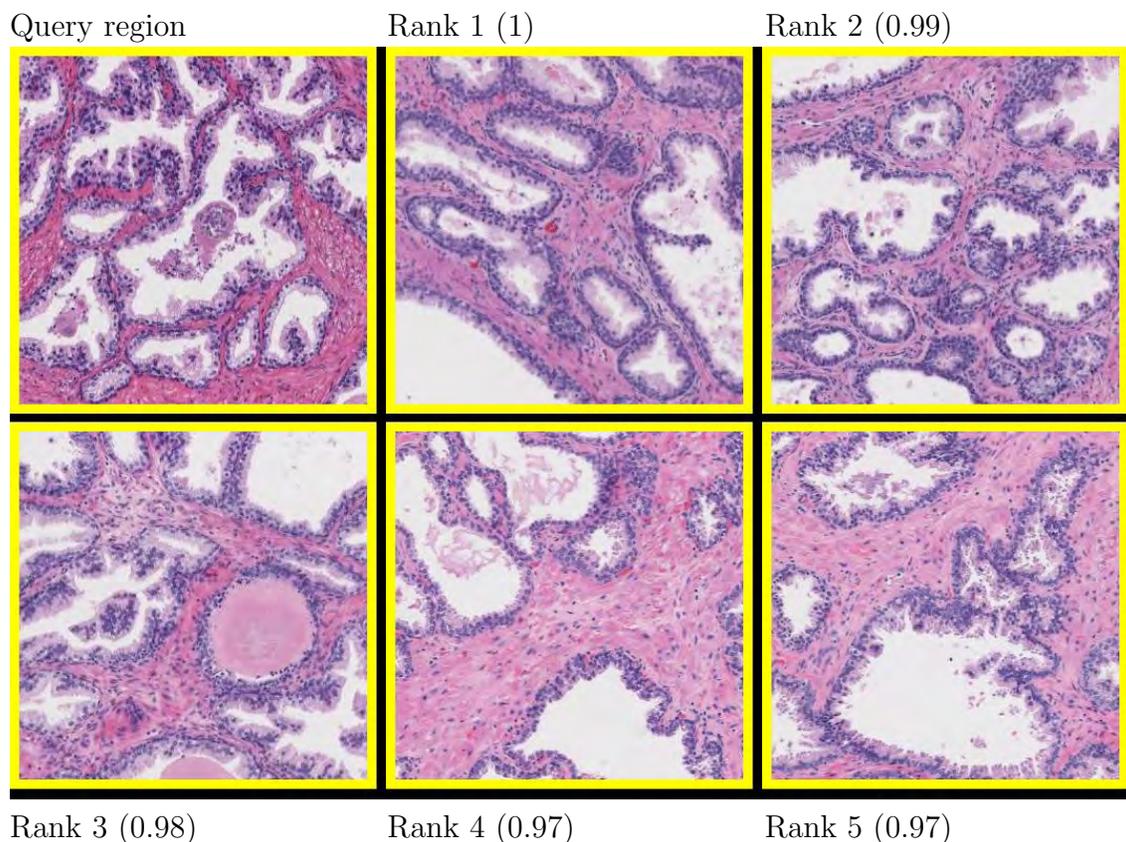
Figure 4.12: A retrieval example, in which the query region is a normal region. The five non-overlapping test regions that have the highest similarity to the query region are shown here. The yellow boundary indicates that the region is labeled as normal. The normalized similarity score to the query region is shown for each retrieved region. In this example, all top five retrieved regions are relevant. The size of all the images is 350×350 pixels, corresponding to a 0.7×0.7 mm^2 tissue region at $5\times$ magnification.

4.6 Summary and Contributions

We have introduced the content-based region retrieval problem in prostate tissue slide images and proposed a gland-based method to address this problem. The gland-based method computes the similarity between two tissue regions based on three attributes: (i) gland density similarity, (ii) gland area similarity, and (iii) gland structure similarity. To compute the gland structure similarity between the two regions, we match the most similar pairs of glands between the two regions, where the similarity between a pair of glands is computed by the reciprocal of the Euclidean distance between the normalized gland features. To evaluate
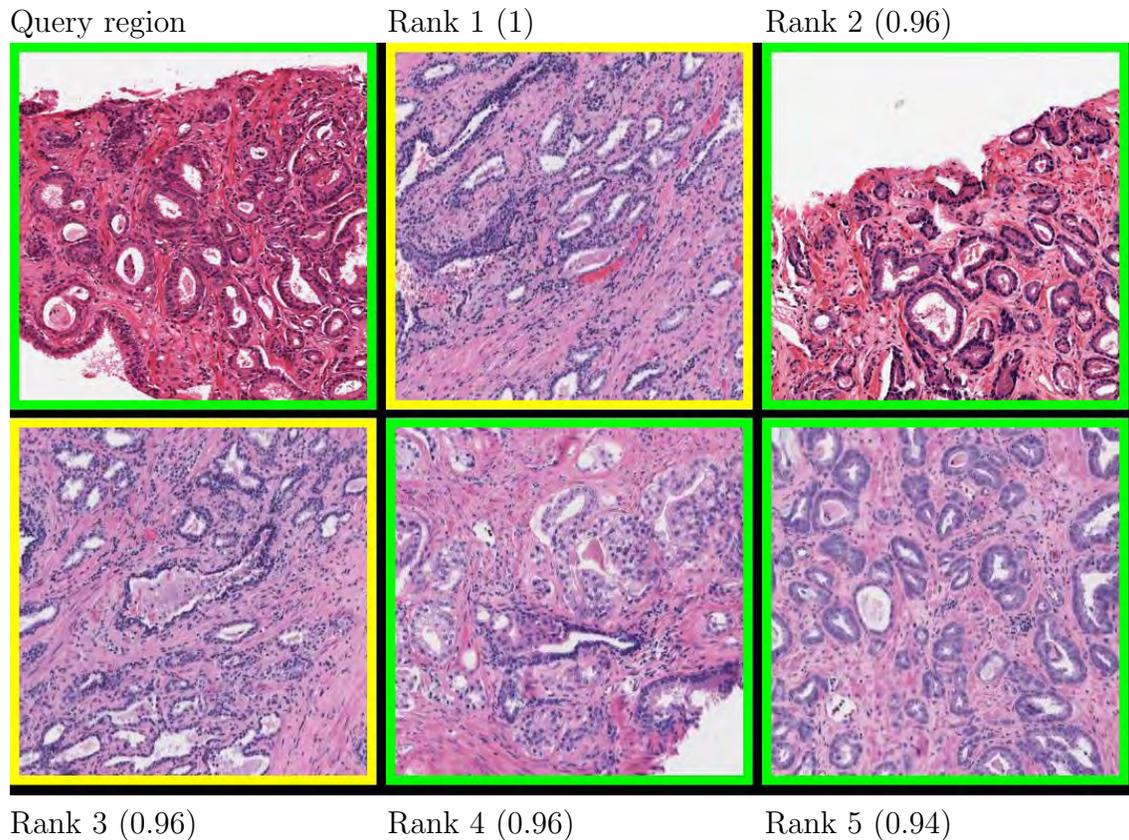
Figure 4.13: A retrieval example, in which the query region is a grade 3 region. The five non-overlapping test regions that have the highest similarity to the query region are shown here. The yellow and green boundaries indicate that the regions are labeled as normal and grade 3, respectively. The normalized similarity score to the query region is shown for each retrieved region. In this example, there are three relevant retrieved regions in the top-5 retrievals. The size of all the images is 350×350 pixels, corresponding to a 0.7×0.7 mm^2 tissue region at $5\times$ magnification.

the proposed method, we compare it with three popular methods in the literature, texture-based, BoW-based and category-based methods. The evaluation metrics include the overall precision-recall relationship and the average number of relevant regions among the top 10 retrieved regions. The experimental results show that the gland-based method performs better than the other three methods for both the metrics. Moreover, we also present a combination of the gland-based and the BoW-based methods and a combination of the gland-based and the category-based methods, which obtain better performance than each of the three methods (BoW-based, category-based, and gland-based) alone.

| Query region | Rank 1 (1) | Rank 2 (0.87) |
| Rank 3 (0.84) | Rank 4 (0.81) | Rank 5 (0.79) |
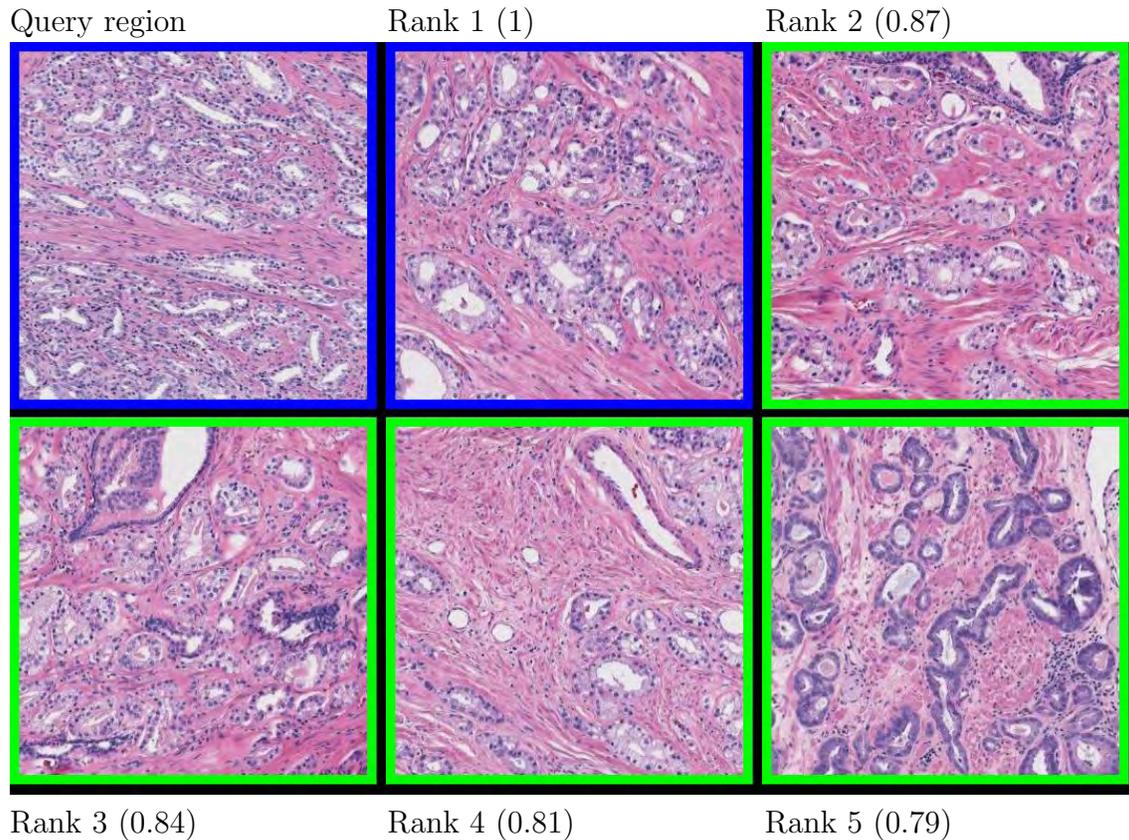
Figure 4.14: A retrieval example, in which the query region is a grade 4 region. The five non-overlapping test regions that have the highest similarity to the query region are shown here. The green and blue boundaries indicate that the regions are labeled as grade 3 and grade 4, respectively. The normalized similarity score to the query region is shown for each retrieved region. In this example, there is only one relevant retrieved region in the top-5 retrievals. The size of all the images is 350×350 pixels, corresponding to a 0.7×0.7 mm^2 tissue region at $5\times$ magnification.

Chapter 5

CYTOLOGICAL FEATURES IN PROSTATE HISTOPATHOLOGY

5.1 Introduction

As mentioned in [107], the Gleason grading method utilizes only structural features of the glands and does not consider cytological features of the tissue. Cytological features include shape, quantity, and arrangement of the basic components of the tissue such as cells, cytoplasm, nuclei, and nucleoli [108, 109]. In contrast to structural features, which do not require high magnification images (in chapter 2, we were able to use $5\times$ images for gland segmentation and classification), cytological features can only be extracted in high magnification images where the details of tissue components can be easily discerned.

The literature contains pathological studies that have used cytological features to improve the diagnostic results [107, 109, 110]. Moreover, we can also see the importance of the cytological features by observing how a pathologist identifies cancer regions in a tissue image. Specifically, the pathologist pays careful attention to nuclei that have prominent nucleoli. Nuclei in cancer regions usually appear light blue and contain nucleoli that appear as small dark spots [111, 112]. In contrast, nuclei in normal tissue regions usually appear uniformly dark or uniformly light over the entire area, with no nucleoli. Figure 5.1 illustrates the difference. This suggests that nuclei with nucleoli provide a significant evidence for the presence of cancer regions in the prostate tissue.

In this chapter, we propose a method to detect the presence of nuclei with nucleoli (NwN), which is one of the most important cytological features for prostate cancer detection. To the best of our knowledge, no study on automatic detection of the NwN for prostate cancer detection and grading has been published. The process of detecting NwN involves segmenting nuclei from the tissue image, and then identifying nucleoli within the nuclei. We

propose three modules to accomplish this task: (i) maximum object likelihood binarization (MOLB) to segment nuclei, (ii) maximum boundary magnitude binarization (MBMB) and (iii) feature-based object identification (FOI) to identify nucleoli in the nuclei. The flowchart of the proposed NwN detection method is shown in Figure 5.2. The remainder of this chapter is organized as follows. Section 5.2 describes the MOLB algorithm and section 5.3 presents the MBMB and the FOI algorithms. Section 5.4 evaluates the NwN detection result by using a prostate cancer detection framework. Finally, section 5.5 presents the conclusions of this chapter.
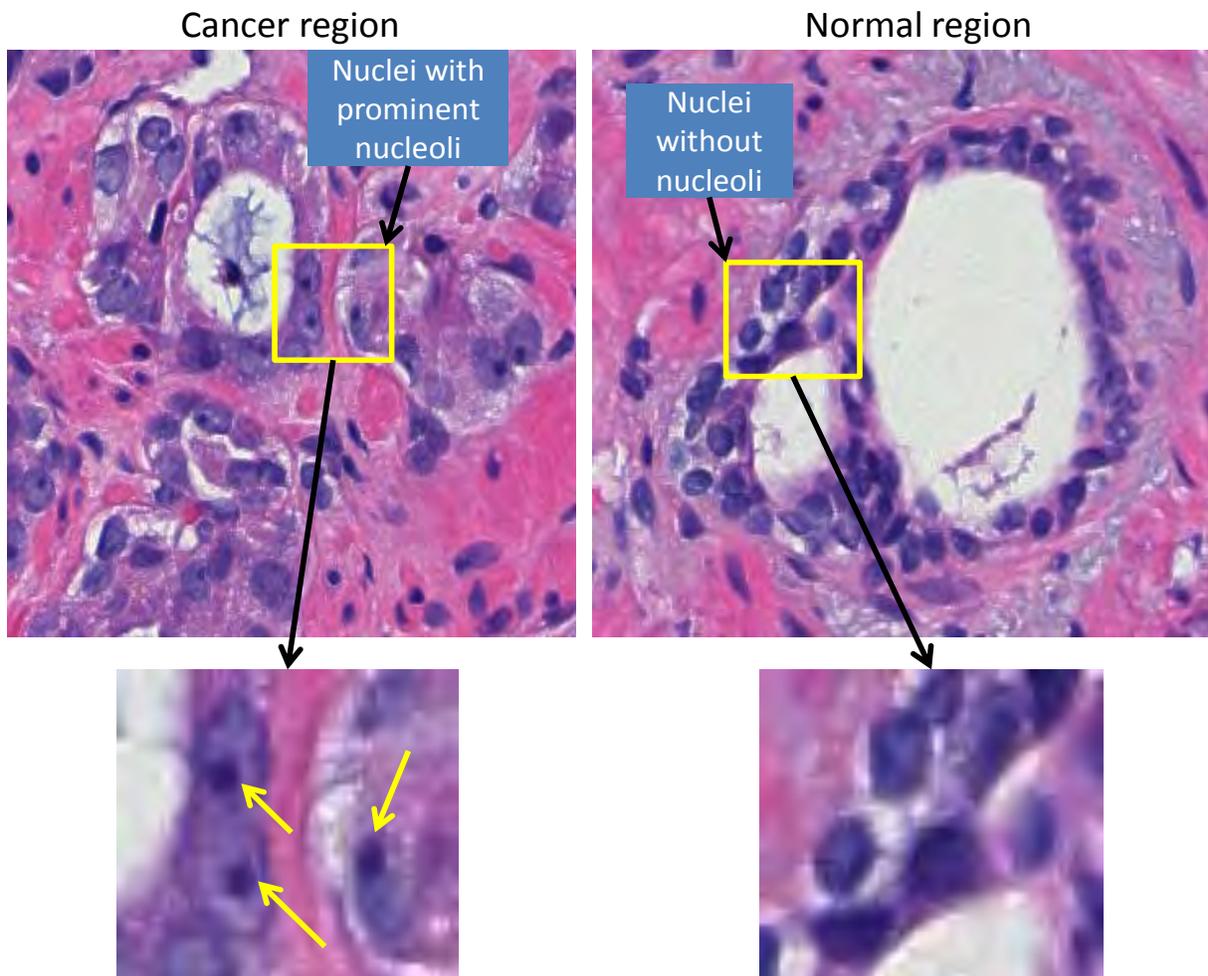
Figure 5.1: Nuclei with prominent nucleoli and nuclei without nucleoli. Nucleoli appear as dark spots inside the nucleus regions (yellow arrows). Nuclei with prominent nucleoli commonly appear in cancer regions. The average image size is 300×280 pixels, corresponding to a 0.15×0.14 mm^2 tissue region digitized at $20\times$ magnification.

Input RGB image

The MOLB algorithm is used to segment all nuclei from the image

For each segmented nucleus

RGB image of the nucleus

The MBMB algorithm is used to segment the dark spots inside the nucleus (denoted by the blank areas inside the nucleus)

Non-nucleolus

Nucleolus

The FOI algorithm is used to identify a nucleolus among all dark spots
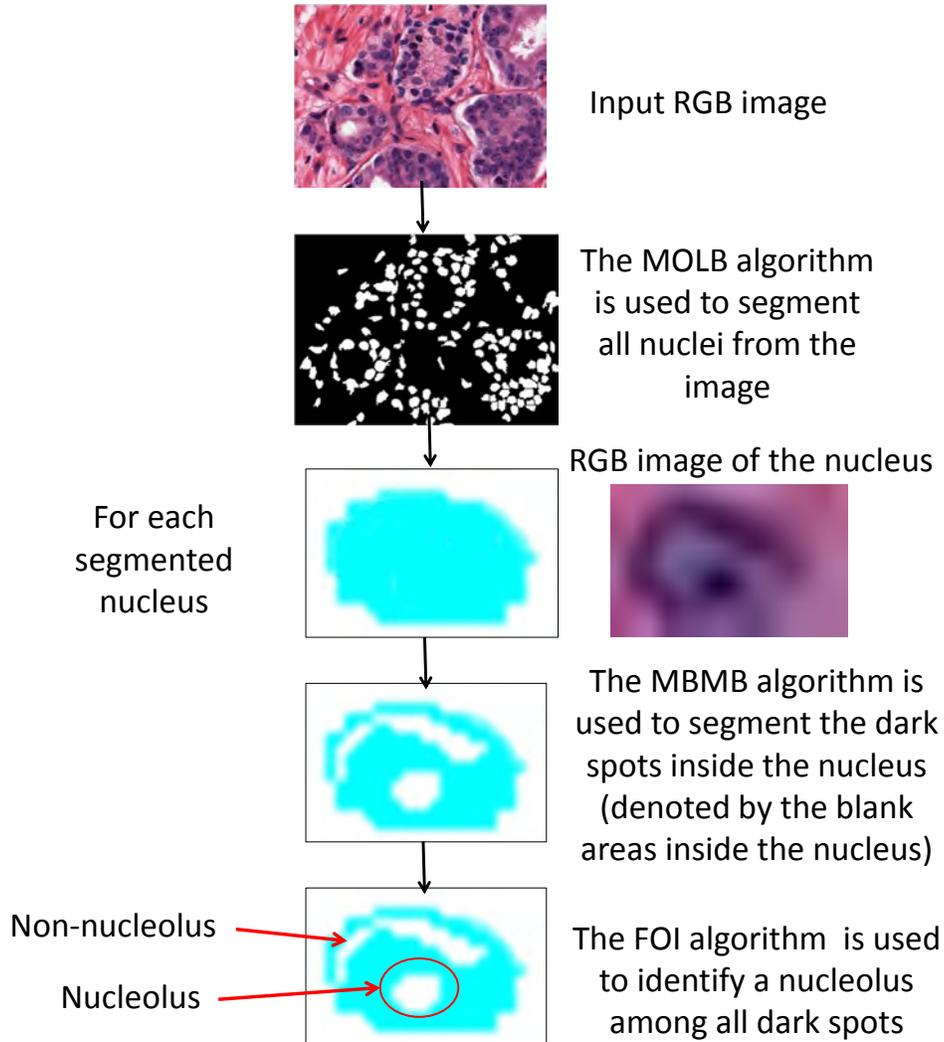
Figure 5.2: Flowchart of the proposed method for detecting nuclei with nucleoli (NwN). The method includes three modules: MOLB (to segment nuclei), MBMB (to segment the dark spots inside the nuclei), and FOI (to identify nucleoli).

5.2 Nucleus Segmentation

5.2.1 Maximum Object Likelihood Binarization (MOLB)

In this problem, we do not use the radial symmetry method mentioned in chapter 3 to segment nuclei because the goal of the radial symmetry method is to detect the nuclei centers but not to segment the entire nucleus region. Hence, to segment nuclei in the tissue image, we propose a *maximum object likelihood binarization* (MOLB) algorithm. In general,

the goal of this algorithm is to segment an object of interest O with the feature vector $f(O)$ in a grayscale image I^1. We first assume that $f(O)$ follows a density g with the parameter vector θ. An estimate $\hat{\theta}$ is obtained from a training set of objects. A threshold t_0 to binarize the grayscale image I is obtained such that the average object likelihood of the foreground blobs is maximized. Formally, let $B_i^t, i = 1, \ldots, n^t$, denote the n^t foreground blobs generated by binarizing I with a threshold $t \in [t_{min}, t_{max}]$ (note that B_i^t and n^t depend on t). We choose t_0 such that:

$$t_0 = argmax_t \frac{1}{n^t} \sum_{i=1}^{n^t} g(f(B_i^t)|\hat{\theta}) \tag{5.1}$$

where $f(B_i^t)$ is the feature vector of blob B_i^t and $g(f(B_i^t)|\hat{\theta})$ is the object likelihood of blob B_i^t since it estimates the similarity of the features for B_i^t and the object of interest O (which have the density g with the parameter $\hat{\theta}$).

In this procedure, after binarizing I using a threshold $t \in [t_{min}, t_{max}]$, we apply a 4-connectivity connected component algorithm to group the foreground pixels (pixels whose intensities are greater than t) into blobs $B_i^t, i = 1, \ldots, n^t$. Next, the blob features, $f(B_i^t)$, and their average object likelihood, $\frac{1}{n^t} \sum_{i=1}^{n^t} g(f(B_i^t)|\hat{\theta})$ are computed. The optimal threshold t_0, computed in Equation 5.1, is the threshold resulting in the maximum average object likelihood. The blobs obtained by binarizing I with t_0 are the outputs of the algorithm. The MOLB algorithm is illustrated in Figure 5.3.

5.2.2 Using the MOLB Algorithm to Segment Nuclei

In order to use the MOLB algorithm to segment nuclei in the prostate tissue image, we first need to obtain a grayscale version of the original RGB tissue image. Since nuclei appear as blue, we transform the RGB color space to the *Lab* color space, and choose the grayscale image as the b channel of this *Lab* color space (we denote this channel as b_{Lab}). As mentioned

---

[1]In the grayscale images shown in this chapter, the darker pixels indicate higher gray values.

Input image

Binary image

Thresholding by t

$t \in [t_{min}, t_{max}]$
$t_{min} = 0.4$
$t_{max} = 0.8$
Step size = 0.05

Grayscale image with
gray values in [0,1]
Darker pixels have higher
gray values

Threshold   Foreground blobs   Avg. object likelihood (L_t)

$$t = t_i \rightarrow \{B_1^i, B_2^i, \ldots, B_{n_i}^i\} \rightarrow \frac{1}{n_i} \sum_{j=1}^{n_i} g(f(B_j^i)|\hat{\theta}) = L_t$$

$$t_0 = argmax_t L_t$$

Thresholding
by t_0, post
processing

Output nuclei
segments
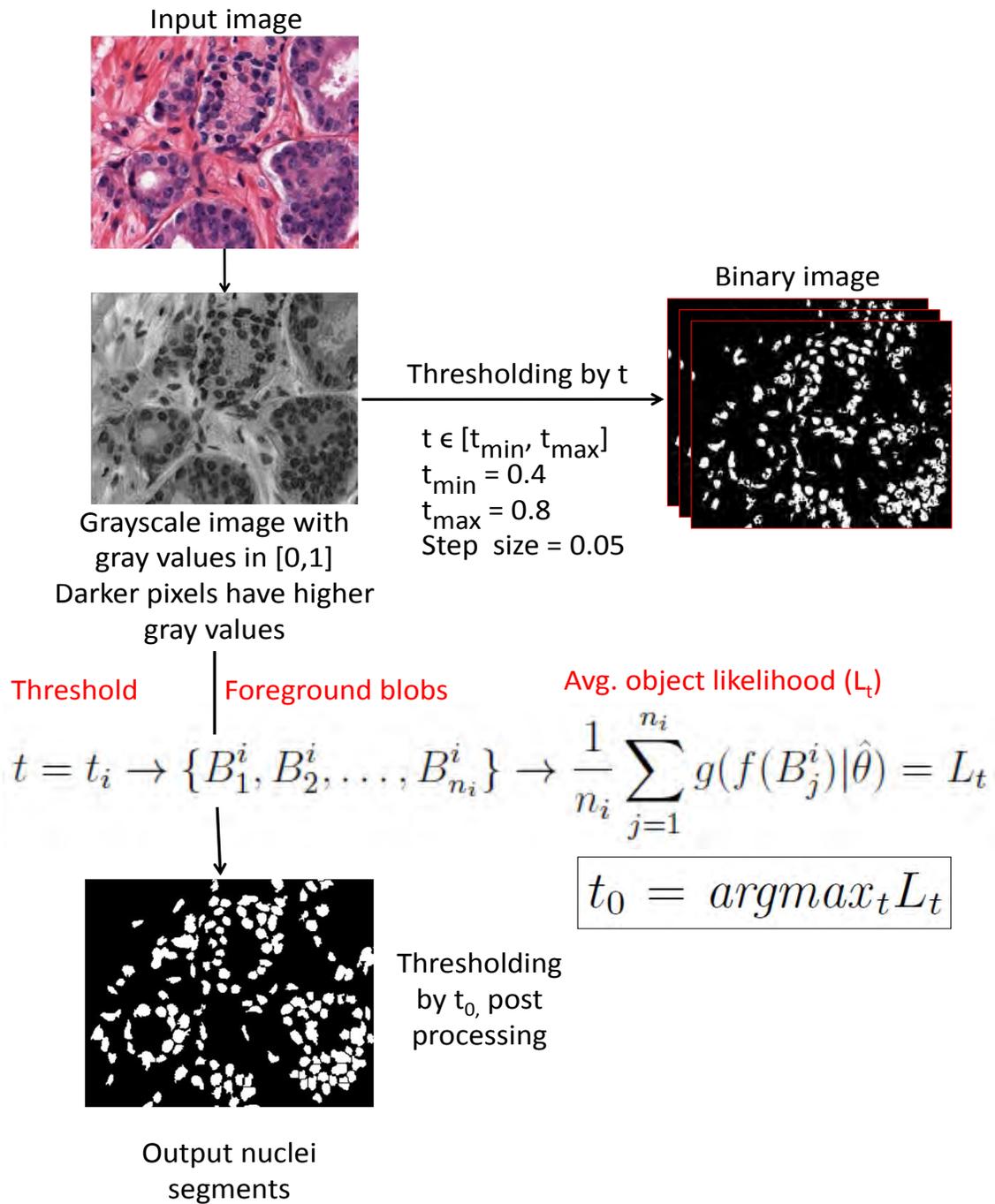
Figure 5.3: The maximum object likelihood binarization (MOLB) algorithm.

149
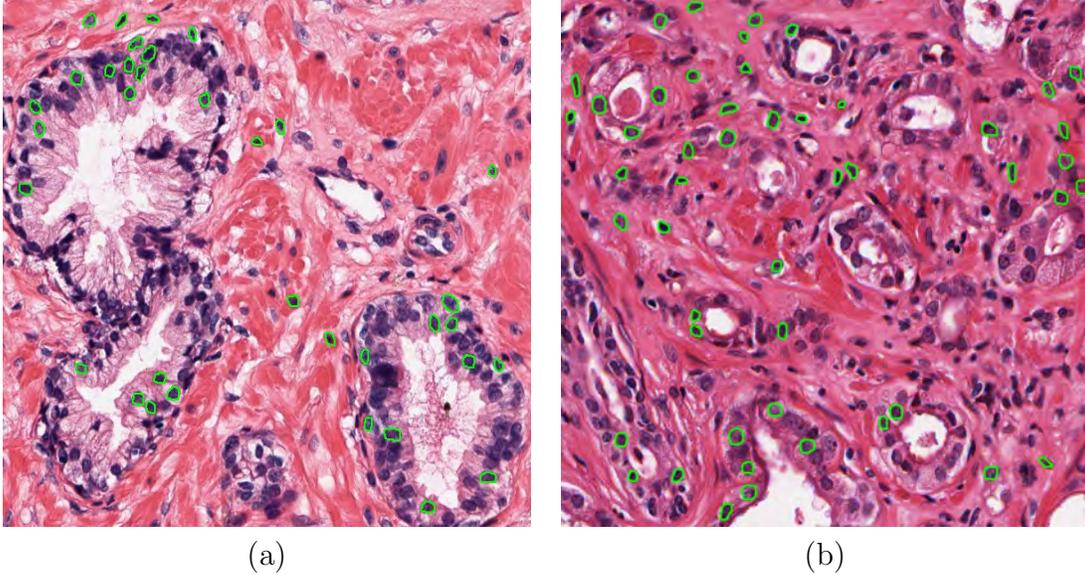
Figure 5.4: Training images for nucleus features used in the MOLB algorithm. The training images are obtained from both normal tissue regions (a) and cancer tissue regions (b), to account for variations in nucleus shape and size between these regions. Manually segmented nuclei are denoted by green contours in these images. The average image size is 600×550 pixels, corresponding to a 0.3×0.27 mm^2 tissue region digitized at $20\times$ magnification.

in chapter 2, the *Lab* color space separates the luminance and chrominance of the colors, so that the b_{Lab} channel represents the blue color. We then normalize the intensity values in the b_{Lab} channel to the range [0,1] to account for color variations among the different tissue slides (some tissue slides may have a darker blue color than others). In Figure 5.5, we qualitatively compare the grayscale images generated by (i) the b_{Lab} channel and (ii) the function *rgb2gray* included in the MATLAB image processing toolbox [113] (we refer to the grayscale image generated by this function as the Matlab-gray image)[2]. As can be seen in Figure 5.5, the b_{Lab} channel shows a stronger contrast between nucleus regions and the background.

The objects O to be segmented by the MOLB algorithm are the nuclei. The feature

---

[2]The grayscale image generated by the function *rgb2gray* only includes luminance information extracted from the RGB image (hue and saturation are discarded). The luminance is extracted by applying the formula Matlab-gray = 0.2989×R + 0.5870×G + 0.1140×B, where R, G, and B denote the 3 channels of the RGB color space. This method is commonly used to generate a grayscale image from an RGB image.
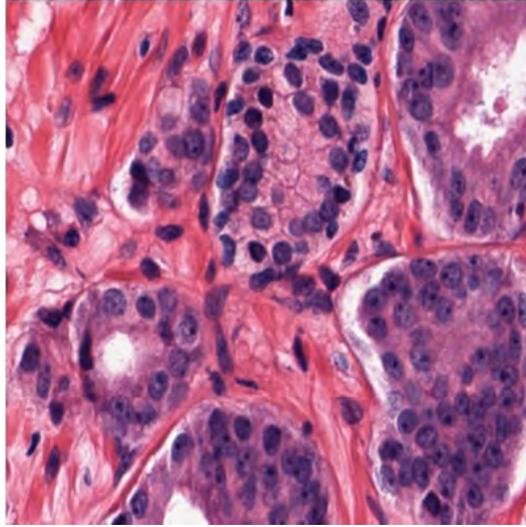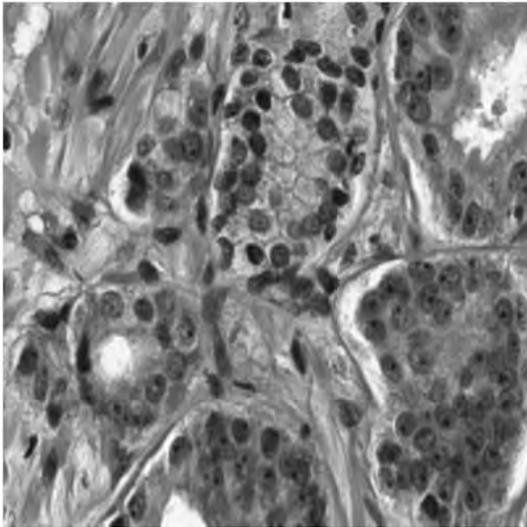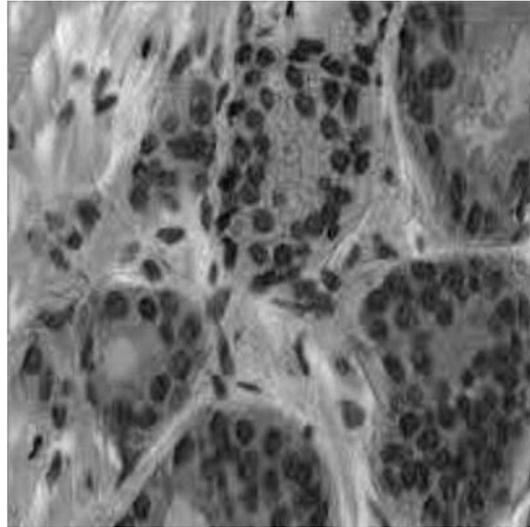
Figure 5.5: The choice of the grayscale image to be used in the MOLB algorithm. (a) Input RGB image. The grayscale images are (b) the Matlab-gray image and (c) the b_{Lab} channel image. The b_{Lab} channel image shows a stronger contrast between nucleus regions and the background. The image size is 400×330 pixels, corresponding to a 0.2×0.16 mm^2 tissue region digitized at $20\times$ magnification.

(a)	(b)
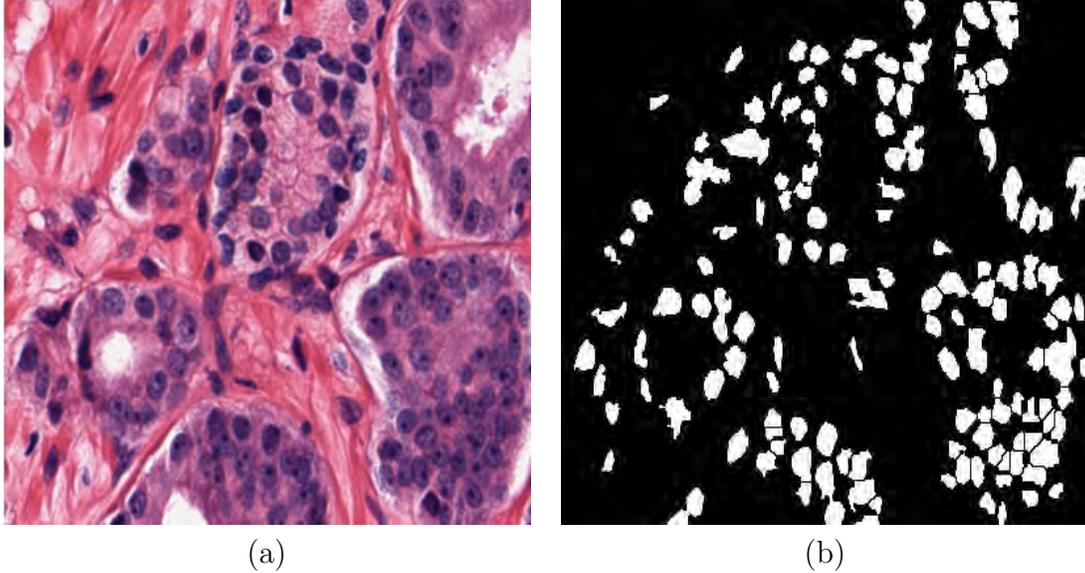
Figure 5.6: Nucleus segmentation result. (a) Input image. (b) Result of nucleus segmentation using the MOLB algorithm. The image size is 400×330 pixels, corresponding to a 0.2×0.16 mm^2 tissue region digitized at $20\times$ magnification.

vector for a nucleus is defined as $f(O) = (a, c)$, where a and c denote the nucleus's area and circularity, respectively. The circularity is defined as $c = (4\pi a)/p^2$, where p is the perimeter of the nucleus. For this problem, we assume that the feature density is bivariate Gaussian $f(O) \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$, where $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are estimated from a training set of manually segmented nuclei (Figure 5.4). Following the binarization procedure, we perform post-processing operations to (i) remove noise and (ii) split connected nuclei by applying the watershed algorithm [114]. An example of the segmentation result is depicted in Figure 5.6.

5.3 Nucleolus Identification

Since nucleoli appear as small dark spots inside the nucleus region (Figure 5.1), we first segment the dark spots from each nucleus. Next, the features of these spots are examined to determine whether a nucleolus is indeed present among the spots. These two operations are performed as follows.

5.3.1 Maximum Boundary Magnitude Binarization (MBMB)

Based on the observation that the boundaries of the dark spots within the nucleus regions are typically salient (i.e. they have a strong gradient magnitude), we propose an algorithm, termed the *maximum boundary magnitude binarization* (MBMB) to segment the dark spots within each nucleus. The MBMB algorithm aims at segmenting the objects with salient boundaries, which are the nucleoli in our case, from a region of interest (ROI), which is the nucleus region in our case.

To perform the segmentation, we need to find a threshold t_0 to binarize the ROI (in a grayscale image) such that a foreground object that has the maximum gradient magnitude on the boundary is generated. Formally, let O_i^t, for $i = 1, \ldots, n^t$, denote the n^t foreground objects obtained in the ROI, R, when R is binarized with a threshold $t \in [t_{min}, t_{max}]$ (note that O_i^t and n^t depend on t). We choose t_0 such that:

$$t_0 = argmax_t(\max_{i \in [1, n^t]} mag(O_i^t)), \tag{5.2}$$

where $mag(O_i^t)$ denotes the average gradient magnitude of the pixels on the boundary of O_i^t.

Similar to the MOLB algorithm, after applying a threshold $t \in [t_{min}, t_{max}]$ to binarize R, we create the foreground objects $O_i^t, i = 1, \ldots, n^t$ in R using the connected component procedure. Given the objects O_i, we compute the average gradient magnitude on the object boundary and find the maximum of the average magnitudes ($\max_{i \in [1, n^t]} mag(O_i^t)$). The best threshold t_0, computed using Equation 5.2, is the threshold that maximizes the object boundary gradient magnitude. This is a local binarization method since in each ROI, we find a different threshold t_0 to binarize and detect the local objects in that ROI.

Unlike nuclei, which are more salient in the b_{Lab} image (as mentioned earlier), the nucleoli are more salient in the Matlab-gray image (Figure 5.8). This is because the b_{Lab} image is better to discriminate between blue pixels (nucleus pixels) and non-blue pixels (background pixels). In contrast, the Matlab-gray image, which contains luminance information, is better to discriminate between dark color pixels (nucleolus pixels) and light color pixels

A segmented nucleus and its grayscale image. The gray values are in [0,1]. Darker pixels have higher gray values

Binary images

Thresholding by t

$t \in [t_{min}, t_{max}]$
$t_{min} = 0.5$
$t_{max} = 0.9$
Step size= 0.05

Threshold

Foreground spots

Avg. boundary magnitude

$$t = t_i \rightarrow \{O_1^i, O_2^i, \ldots, O_{n^i}^i\} \rightarrow \{mag(O_1^i), mag(O_2^i), \ldots, mag(O_{n^i}^i)\} \rightarrow$$

$$\max_{j \in [1, n^i]} mag(O_j^i) = M_t$$

Max of the avg. magnitude

$$t_0 = argmax_t M_t$$

Thresholding by t_0
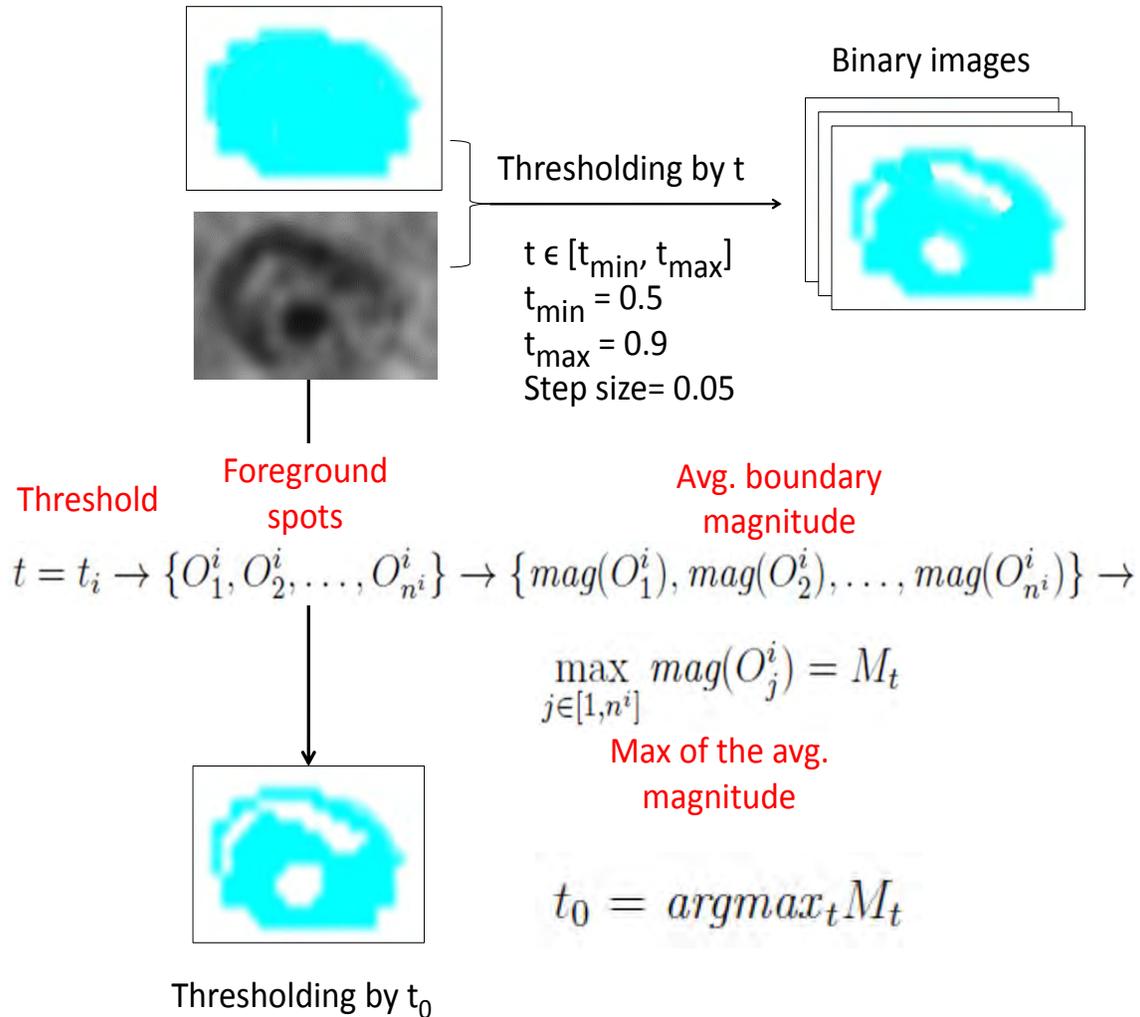
Figure 5.7: The maximum boundary magnitude binarization (MBMB) algorithm.

(the remaining areas in the nucleus regions). Therefore, we apply the MBMB algorithm to the Matlab-gray image to segment the nucleoli. However, this may result in several dark spots within each nucleus because noisy dark regions resulting from a poor tissue staining procedure may also generate spots. Figure 5.7 depicts the MBMB algorithm.

<div style="text-align:center">(a)   (b)   (c)</div>
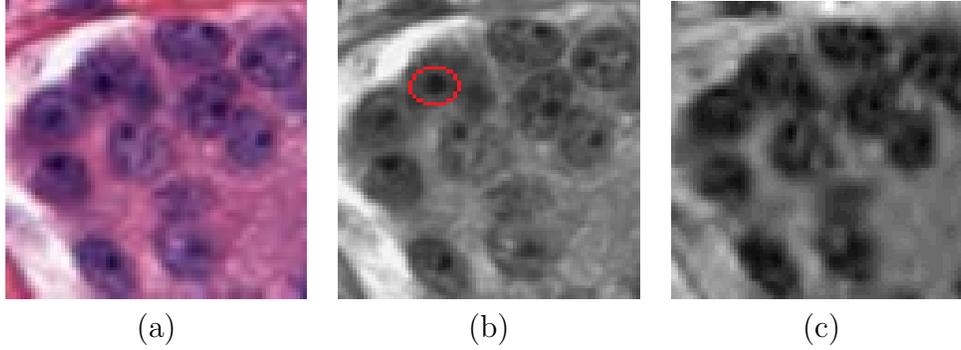
Figure 5.8: The choice of grayscale image to be used in the MBMB algorithm. The nucleoli (the dark spot enclosed by the red circle in (b)) in the RGB image (a) appears more salient in the Matlab-gray image (b) than in the b_{Lab} channel image (c). The image size is 80×90 pixels, corresponding to a 0.04×0.045 mm^2 tissue region digitized at $20\times$ magnification.

5.3.2 Feature-based Object Identification (FOI)

We need to identify an object of interest O^* (nucleolus) among a pool of objects $\{O_i\}_{i=1}^n$ (dark spots)[3]. Similar to the formulation of the MOLB algorithm in section 5.2, we associate the object of interest O^* with a feature vector $f(O^*)$. We also assume that $f(O^*)$ follows a density g with the parameter vector θ. An estimate $\hat{\theta}$ is also obtained from a training set of manually segmented nucleoli (Figure 5.9).

For each object O_i, we compute its feature vector $f(O_i)$. Next, we choose the object O_m which has the maximum likelihood: $O_m = argmax_i g(f(O_i)|\hat{\theta})$ (this is the most similar object to O^* among all the objects). The object O_m is considered the object of interest if $f_j^{min} < f_j(O_m) < f_j^{max} \ \forall j$, where the constraints for each feature, i.e. f_j^{min} and f_j^{max} are estimated from the training set. We again use area and circularity as the features to identify the nucleoli. By using this algorithm, the nuclei in which nucleoli are found are considered NwN. Figure 5.10 shows an example of the NwN detection result.

---

[3]Although there may be more than one nucleolus in a nucleus, we only need to identify one of them.

<div style="text-align:center">155</div>
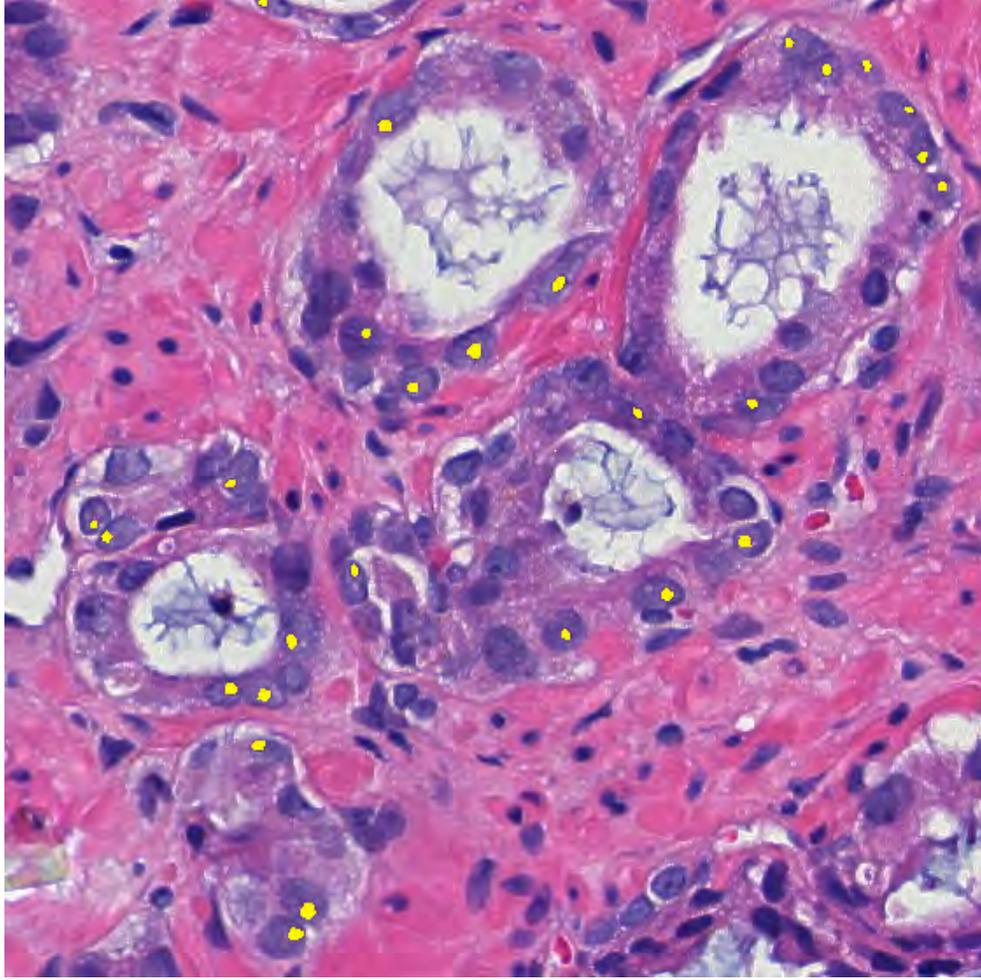
Figure 5.9: Manually segmented nucleoli (highlighted in yellow), whose features are used in the FOI algorithm. The image size is 560×530 pixels, corresponding to a 0.28×0.26 mm^2 tissue region digitized at $20\times$ magnification.

5.4 Evaluation of NwN Detection Result

Determining whether nucleoli are contained in a nucleus is a laborious and time-consuming task for pathologists. This explains the lack of ground truth for the NwN in the images. Therefore, we evaluate the NwN detection results via (i) the prostate cancer detection results and (ii) the tissue image classification results.
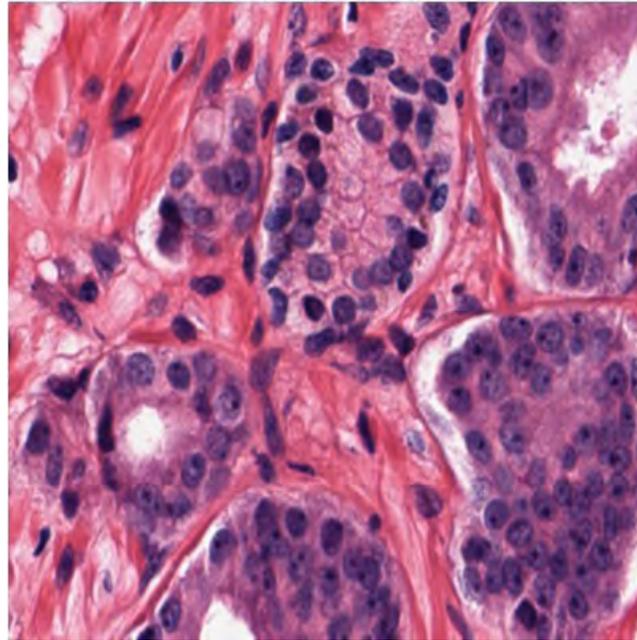
(a)
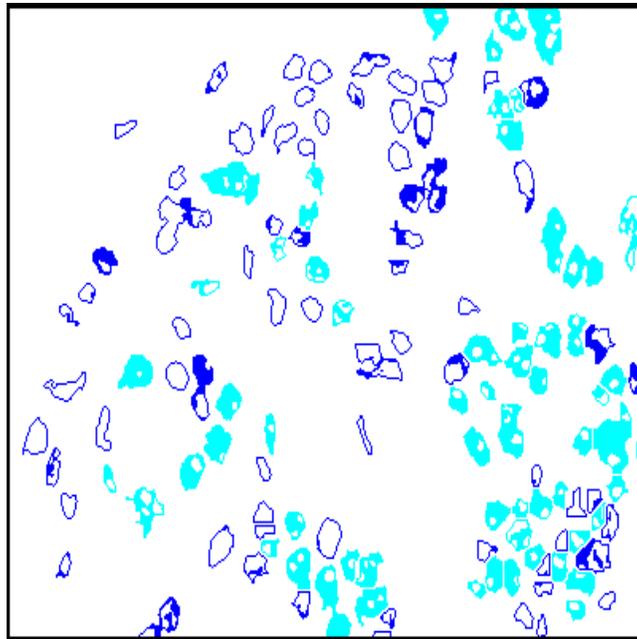
(b)

Figure 5.10: NwN detection result. (a) Input image. (b) NwN detection result. Cyan regions denote NwN and blue regions are nuclei without nucleoli. The blank areas inside each nucleus region are the segmented dark spots obtained by the MBMB algorithm. In the cyan regions, these spots are round and small in size and are considered to be nucleoli. In the blue regions, these spots are either elongated or much larger in size and are not considered to be nucleoli. The image size is 400×330 pixels, corresponding to a 0.2×0.16 mm^2 tissue region digitized at $20\times$ magnification.

157

5.4.1 Evaluation by Prostate Cancer Detection Results

In this evaluation method, we utilize the pathologist's annotations for cancer regions in each image (Figure 5.11) as the ground truth for quantitatively evaluating the proposed NwN detection method. In other words, we use the NwN detection result to identify cancer regions in the prostate tissue, and then we compare these regions with the ground truth. By doing this, the cancer detection results reflect the performance of the proposed NwN detection method. Moreover, by using information about the NwN (a cytological feature), we also show that cytological features are useful for the automated systems that perform prostate cancer detection. For generality, we will refer to the NwN information as the cytological feature in the remainder of this section.

For the prostate cancer detection problem, we employ the patch-based detection approach, which was previously mentioned in [45][4]. We divide the image into a grid of patches and classify each patch as a normal patch or a cancer patch, based on the features extracted for that patch. To demonstrate the robustness of the cytological feature (and the NwN detection method), we also use textural features for cancer detection and compare them with the cytological feature. The textural features and the prostate cancer detection algorithm are described in the following subsections.

5.4.1.1 Textural Features

Similar to [45], the textural features computed for an image patch include the first-order statistics, second-order statistics, and Gabor filter features. There are four first-order statistic features comprising mean, standard deviation, median, and gradient magnitude of pixel intensity in the patch. For the second-order statistic features, we form the gray level co-occurrence matrix for all the pixels in the patch and compute 13 features from this matrix [4] which include energy, correlation, inertia, entropy, inverse difference moment, sum average,

---

[4]Doyle et al. [45] showed that the patch-based approach was more robust than the pixel-based approach that they employed.
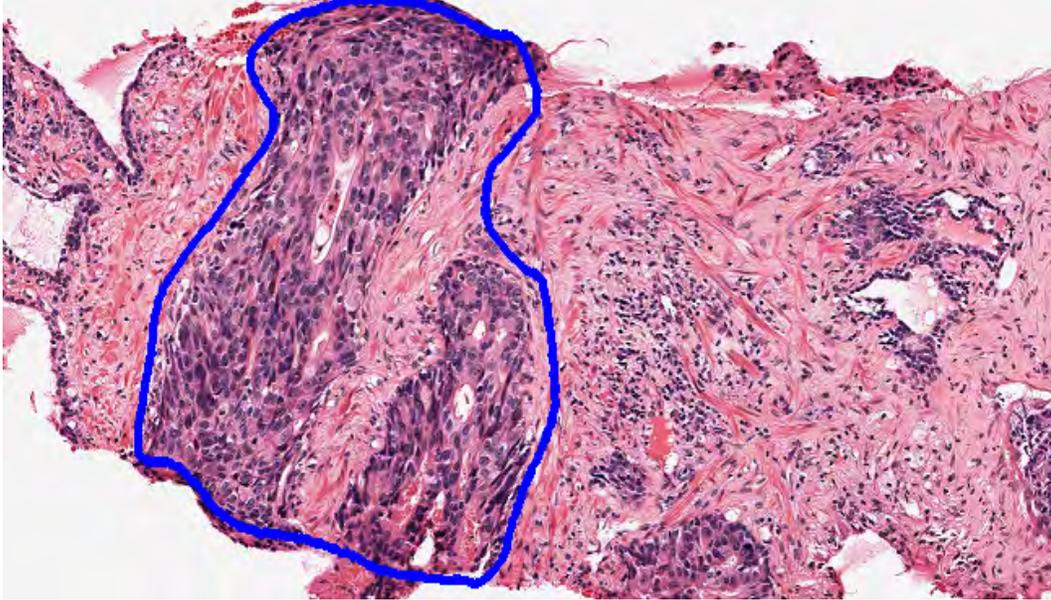
Figure 5.11: Cancer annotation performed by a pathologist, which is used as the ground truth for evaluating the NwN detection method. The region inside the blue contour is annotated as a cancer region. Regions outside the contour are considered normal regions. The image size is 1,300 × 2,600 pixels, corresponding to a 0.65 × 1.3 mm^2 tissue region digitized at 20× magnification.

sum variance, sum entropy, difference average, difference variance, difference entropy, and two information measures of correlation. For the Gabor filter based texture features [99], we create a bank of 10 filters by using two different scales and five different orientations. The mean and variance of the filter response are used as features. Thus, a total of 20 features are extracted by using Gabor filters. We obtain 37 features using these three feature types (4 first-order statistics features, 13 second-order statistics features and 20 Gabor features). By considering texture in each of the three normalized channels of the *Lab* color space of the image separately, we have a total of $3 \times 37 = 111$ textural features for each patch.

5.4.1.2 Cancer Detection Algorithm

A grid of patches, each consisting of $S \times S$ pixels, is superimposed on the image. In general, we assume that n feature sets are extracted from each image patch. Let $\{\boldsymbol{x}_i\}_{i=1}^{n}$ denote the n feature sets associated with a patch P. Each set $\boldsymbol{x}_i$ may contain one or more features. We

train n classifiers $\{f_i\}_{i=1}^n$, one for each feature set, using a training set of patches and their labels (normal patch or cancer patch). A patch P is classified as a cancer patch if:

$$\prod_i^n P(f_i(\boldsymbol{x}_i) = 1) > \prod_i^n P(f_i(\boldsymbol{x}_i) = 0), \tag{5.3}$$

where $P(f_i(\boldsymbol{x}_i) = 0)$ and $P(f_i(\boldsymbol{x}_i) = 1)$ denote the probabilities that the classifier f_i classifies the patch (using the feature set $\boldsymbol{x}_i$) as a normal patch or a cancer patch, respectively. Otherwise, P is considered a normal patch.

Since we have two feature sets associated with each patch, the cytological feature set (which includes a single feature, namely, *the number of pixels in the patch belonging to NwN*) and the textural feature set (which includes 111 textural features, as described above), we can define three different cancer detection methods:

1. A texture-based method, in which we only use the textural feature set to classify the patch (i.e. the number of feature sets used in Equation 5.3 is 1, $n = 1$),

2. A cytological-feature-based method, in which we only use the cytological feature set to classify the patch ($n = 1$), and

3. A feature fusion method, in which we use both the cytological feature set and the textural feature set to classify the patch ($n = 2$).

All the classifiers f_i used in these three methods are support vector machines (SVM) with a radial basis function (RBF) kernel, each with a different feature set. By comparing the performance of these three cancer detection methods, we can demonstrate the usefulness of the cytological feature, as well as the robustness of the NwN detection method.

The grid size and placement are chosen based on [45], where the authors superimposed the image with a uniform grid that divides the image into 30×30 regions. Similarly, we divide the image into 40×40 patches (that is, $S = 40$) and perform patch classification.

Once cancer patches have been identified, we create contiguous cancer regions by grouping together neighboring cancer patches, where each group O is a set of cancer patches $\{P_i\}_{i=1}^m$

160

such that $\forall P_i \in O, \exists P_j \in O$ with $d(P_i, P_j) \leq t_d$. For each group, we create one contiguous cancer region by generating a convex hull of all the patches; all the pixels inside this convex hull are labeled as cancer pixels. Malignant tumors are characterized by uncontrolled growth, which makes the spread and shape of the tumors extremely difficult to model. We, therefore, make a reasonable simplification, which is to compute the convex hull that includes all the cancer regions that may correspond to a single tumor. The grouping process mimics the pathologist's annotation strategy. We can observe in the ground truth that the pathologist prefers to annotate large cancer regions that include several neighboring cancer glands, rather than individual glands.
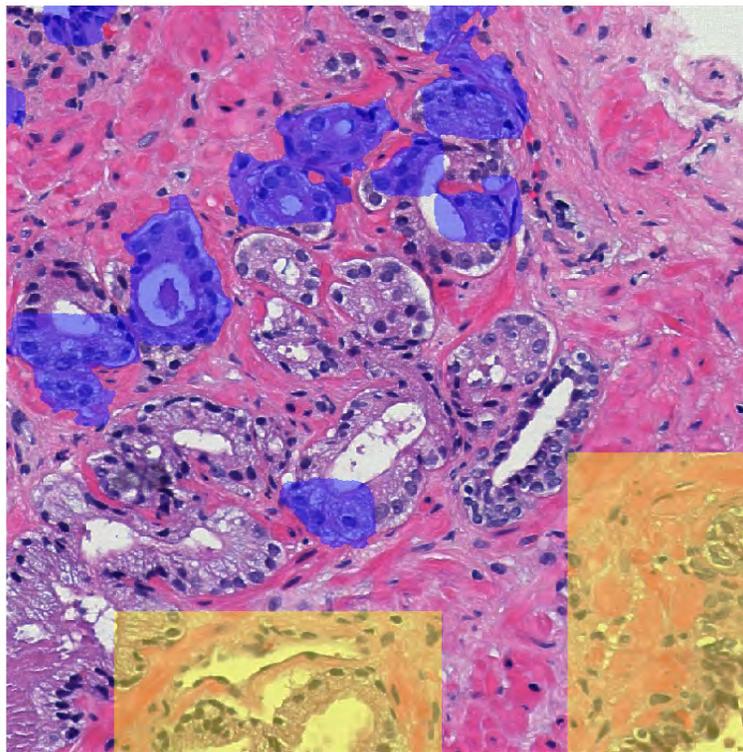
Figure 5.12: A subimage of a training image in which all cancer glands are highlighted in blue and selected normal regions are highlighted in yellow. The image size is 700×770 pixels, corresponding to a 0.35×0.38 mm^2 tissue region digitized at $20\times$ magnification.
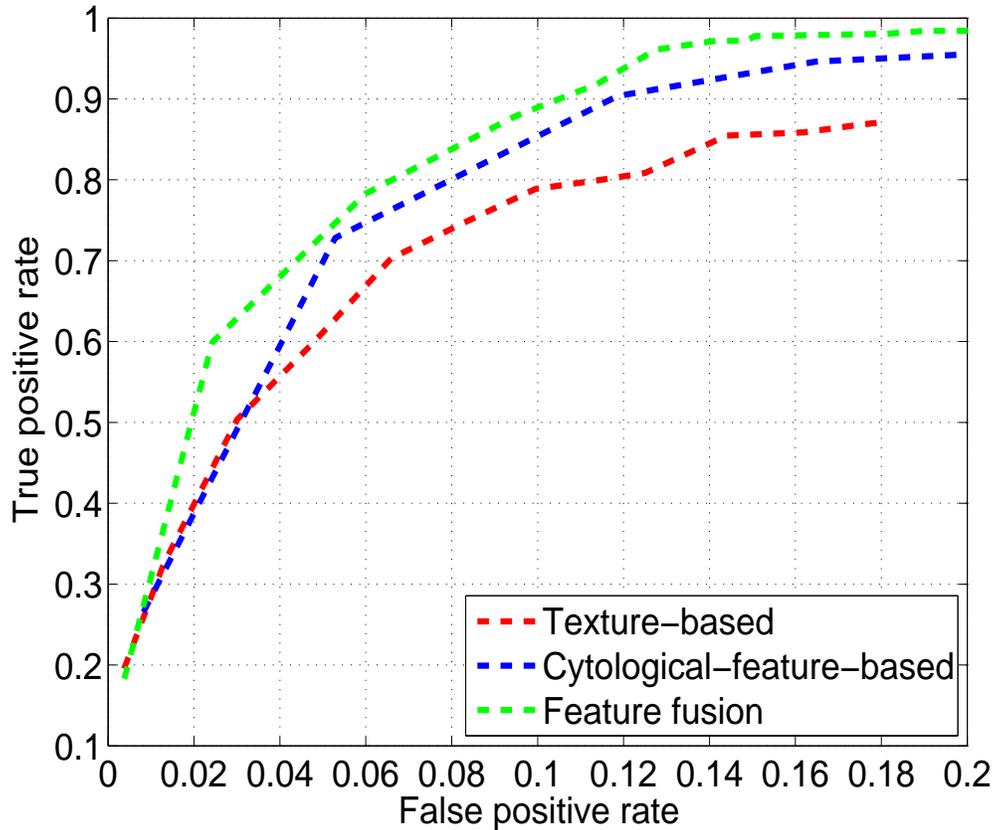
Figure 5.13: ROC curves, which plot the average FPR against the average TPR over all test images, for the feature fusion method, the texture-based method, and the cytological-feature-based method, obtained by varying the threshold t_d from 0 pixels to 500 pixels.

5.4.1.3 Cancer Detection Results

As mentioned in chapter 1, the database used in this experiment includes 17 tissue slide images at $20\times$ magnification. This database contains independent training and test sets. The training set includes 6 images (approximate size is $3{,}100\times3{,}700$ pixels at $20\times$ magnification). In each training image, all the cancer glands were marked by a pathologist. In the remaining non-cancer area, we manually selected a number of normal regions which contain various benign structures of the tissue (stroma, normal glands with different sizes, nuclei). Figure 5.12 shows a portion of a training image. The independent test set consists of 11 tissue images at $20\times$ magnification (approximate size is $3{,}600\times13{,}600$ pixels). The difference in size of the training and testing images is not an important issue because what we need for
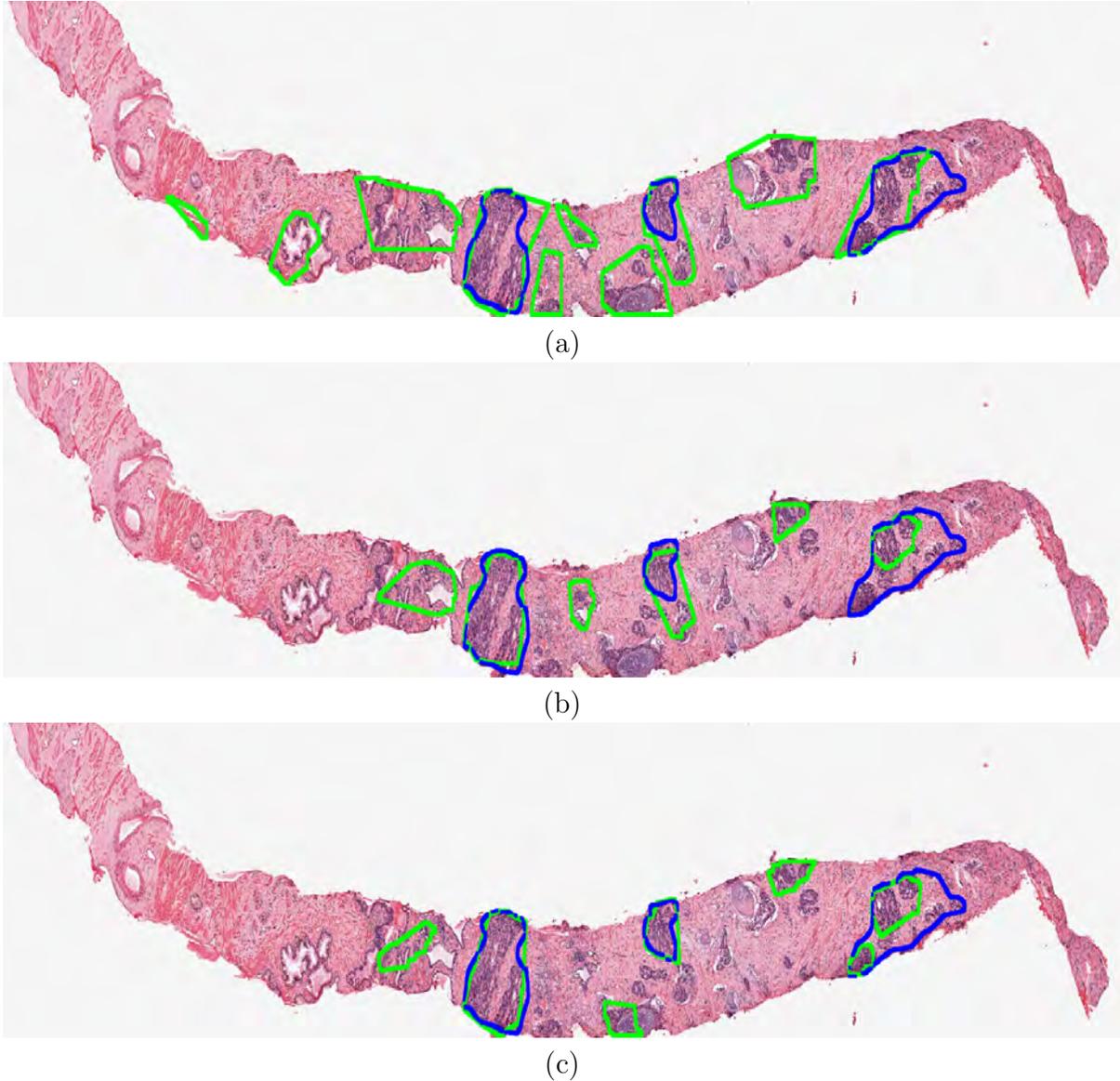
Figure 5.14: Cancer detection results for a test image. The blue contours depict the cancer region annotated by the pathologist, and the green contours depict the results of the algorithm. (a) Result of the texture-based method; (b) result of the cytological-feature-based method and (c) result of the feature fusion method. The threshold $t_d = 90$ pixels was used for all three methods. The image size is $4,000 \times 18,000$ pixels, corresponding to a 2×9 mm^2 tissue slide digitized at $20\times$ magnification.

training is the local information about the patches but not the global information about the entire image. The ground truth for the test images (all cancer regions) were also manually labeled by a pathologist. All the non-labeled regions are considered normal. While in the ground truth of the test set, the pathologist annotated the entire cancer region (including
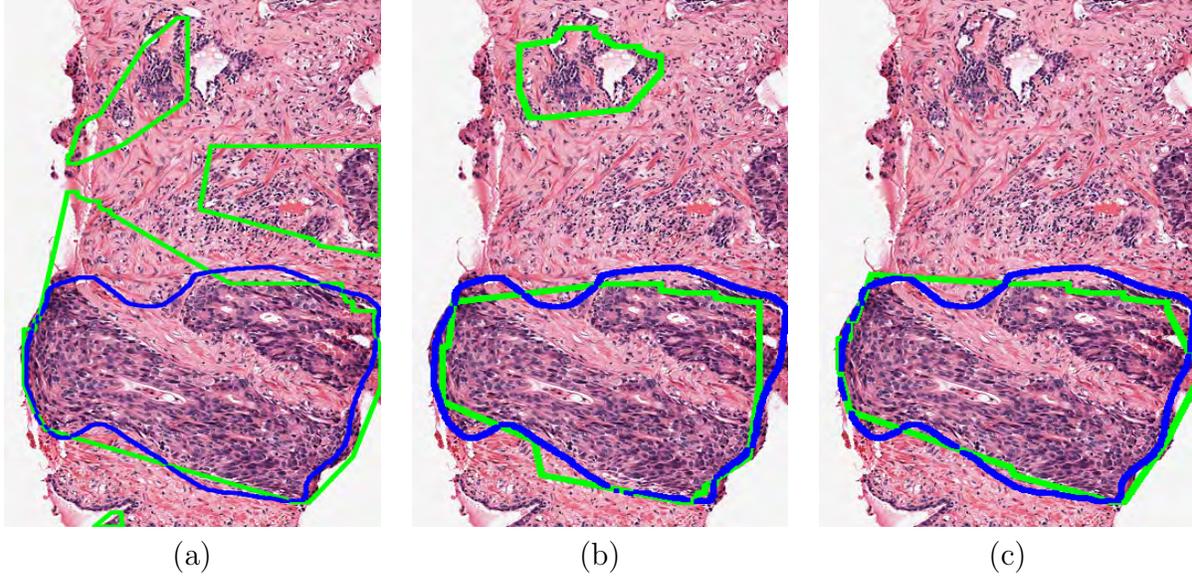
Figure 5.15: Detection results for texture-based method (a), cytological-feature-based method (b), the feature fusion method (c) for a close-up region sampled from the image in Figure 5.14. The image size is 690 × 1,150 pixels, corresponding to a 0.34 × 0.57 mm^2 tissue region digitized at 20× magnification.

neighboring cancer glands and the intervening normal structures such as stroma), only pixels belonging to the cancer glands are annotated as cancer in the training set. In other words, the training data are more reliable.

To evaluate the detection results of the three methods, i.e. cytological-feature-based, texture-based and feature fusion, we compute the true positive rate TPR = TP/(TP + FN) and the false positive rate FPR = FP/(TN + FP), where TP, FP, TN, and FN denote the true positive, false positive, true negative, and false negative, respectively, for a test image. The TPR and FPR are then averaged over all test images. The value of the threshold t_d has a significant influence on both TPR and FPR. When t_d increases, both TPR and FPR increase, i.e. there are more true cancer regions being detected while at the same time, more normal regions get incorrectly classified as cancer regions. Hence, to analyze the relationship between TPR and FPR, we compute these two measures for a wide range of t_d values, [0, 500] pixels, and plot the receiver operating characteristic (ROC) curve in Figure 5.13. In this figure, for the same FPR, the feature fusion method obtains the highest TPR, followed

by the cytological-feature-based method, while the texture-based method obtains the lowest TPR.

Figure 5.14 shows the detection results of the three methods on a tissue image when $t_d = 90$ pixels and Figure 5.15 shows a close-up region of the same image for better details. For this image, all the three methods can find most of the cancer regions. However, the feature fusion method leads to the least number of false detections, followed by the cytological-feature-based method, while the texture-based method has the most false detections. From these results, we can see that the cytological feature boosts the cancer detection accuracy, which suggests that the NwN are well-detected by the proposed method.

5.4.2 Evaluation by Tissue Image Classification Results

In this evaluation method, we reconsider the tissue image classification problem discussed in chapter 3, i.e., classifying a prostate tissue image into normal, Gleason grade 3 and Gleason grade 4. The same database used in chapter 3, including 317 tissue images (113 normal, 134 grade 3, and 70 grade 4), is used here. For each image, we detect the NwN and compute the ratio of the number of NwN to the total number of nuclei in the image, denoted by r_N. Hence, r_N can be considered as the cytological feature of the image.

Once r_N is computed for all the images in the database, we obtain the average values of r_N for normal, grade 3 and grade 4 images as 0.14, 0.21, and 0.20, respectively, which means there are more NwN detected in cancer images (grade 3 and grade 4) compared to normal images. This result demonstrates the effectiveness of the proposed NwN detection method. We further combine this cytological feature with the glandular features (which contain 22 features as discussed in chapter 2) to form a 23-dimensional feature vector to perform the tissue image classification task. We conduct an experiment similar to the tissue image classification experiment in chapter 3 (using the 10-fold cross validation technique and an SVM classifier with RBF kernel) and obtain the following accuracies (with standard deviations) for the normal vs grade 3, normal vs grade 4 and grade 3 vs grade 4 classification

problems: 97.4% (4.5), 99.1% (2.2) and 83.0% (6.5), respectively (the corresponding accuracies obtained using the glandular features alone are 96.7% (4.1), 98.3% (2.3), and 83.3% (7.1), respectively). It can be seen that the classification accuracies for normal vs grade 3 and normal vs grade 4 are slightly improved while the classification accuracy for grade 3 vs grade 4 is not. Since both grade 3 and grade 4 images are cancer images, the NwN information cannot be used to discriminate them. For normal vs cancer (grade 3 or grade 4) classification, since the glandular features are very distinctive and are the dominant features (glands appear very clearly in normal images and are very different from glands in cancer images), using the cytological feature does not improve the classification result dramatically.

In summary, although the ground truth for NwN in the images is not available, we are still able to evaluate the proposed NwN detection method by conducting experiments on the two important problems in prostate cancer diagnosis, namely prostate cancer detection and tissue image classification.

5.5 Summary and Contributions

Although the Gleason grading method is widely used to grade prostate cancer, it only focuses on structural information about the glands and does not utilize the cytological features of the tissue, such as information about the nuclei. However, these cytological features also provide helpful information for detecting cancer in the prostate tissue, and so it is useful to explore cytological features in computer-aided diagnosis systems. In this chapter, we have explored one of the most important cytological features for prostate cancer detection, namely the presence of NwN. This feature has not previously been exploited in any computer aided system for prostate cancer diagnosis.

We have developed a method to detect NwN in the tissue image. The NwN detection procedure includes segmenting nuclei, segmenting the dark spots inside each nucleus, and identifying a nucleolus among these spots. Since we do not have the ground truth for NwN in the images, we evaluate the proposed NwN detection method by performing the prostate

cancer detection problem and the tissue image classification problem. In the tissue image classification problem, we show that NwN appear more commonly in cancer images (Gleason grade 3 and grade 4) than in normal images. In the prostate cancer detection problem, to show the robustness of the NwN detection method, we compare the cancer detection result using the detected NwN with that of a texture-based method. Moreover, the detection result of the proposed feature fusion method is also presented. Through these evaluations, we have shown that the proposed method for NwN detection is robust, and we have also demonstrated that the use of cytological features, in this case the presence of NwN, is helpful in computer-aided diagnosis systems.

Chapter 6

SUMMARY AND FUTURE WORK

6.1 Summary and Contributions

Prostate cancer is a severe threat to men's health and lives. Among the various procedures used in prostate cancer diagnosis, the Gleason grading of the tissue slide taken from the prostate biopsy is the most important procedure. Due to the large volume of the prostate tissue image data being generated, the inconsistency among the pathologists in grading the tissue, the high cost and the low throughput of the grading process, there is a need for computer-aided diagnosis (CAD) systems that can assist pathologists in the Gleason grading task. In this thesis, we have made the following contributions to the development of such a CAD system.

1. A lumen-based gland segmentation method, termed nuclei-lumen-association (NLA), that considers lumen as the central component of the gland, and find nuclei associated with the lumen. The NLA method leads to better gland classification results than the level set method, and comparable gland classification results to the manual gland segmentation method.

2. The gland features that capture rich information about the gland, including structural and contextual information. We show that these proposed gland features are (i) better than gland features used in the published studies in solving the gland classification problem and (ii) better than texture features used in the published studies in solving the tissue image classification problem.

3. A nuclei-based gland segmentation method that models the relationship between nuclei and lumina in the image by a nuclei-lumina-graph. The edges in the graph correspond

to the links between nuclei and nuclei and between nuclei and lumina. The recursive normalized cut method is applied to partition the graph into different components, each of which corresponds to a gland. The nuclei-based method overcomes the major limitation of the conventional lumen-based methods: it is able to detect glands without lumen and glands with multiple lumina.

4. A method to compute the gland-score of a segmented region that measures how similar the segmented region is to a gland unit. The gland-scores are combined with the structural-contextual features to improve the grade 3 vs grade 4 tissue image classification result.

5. Both the lumen-based and nuclei-based methods are evaluated on both low magnification ($5\times$) and high magnification images ($20\times$) to evaluate the effect of image magnification on the performance of the methods.

6. A gland-based method to compute the similarity between two tissue regions. This method can be used to search for regions similar to a region of interest (ROI) in the annotated tissue slides. The retrieved regions can serve as the references that a medical student or a technician can rely upon when grading the ROI.

7. A method to extract an important cytological feature in the prostate tissue images, i.e. the presence of nuclei with prominent nucleoli. To our knowledge, this cytological feature has not yet been exploited in any CAD system. We also demonstrate the usefulness of this cytological feature by applying it in a prostate cancer detection framework.

6.2   Future Work

We notice a large variation in the color of various tissue components in histopathological prostate tissue images. Hence, it would be useful to develop a color normalization method

for the tissue images. This normalization procedure can help to improve the identification of the tissue components, which relies on the color information.

To improve the nuclei-based gland segmentation method, we need to improve the stroma detection result. Besides color information, other types of information such as texture may be also considered.

Additional work needs to be done in the identification of Gleason grade 4 images since these images commonly contain glands with nuclei-forming closed chain structure similar to glands in grade 3 images.

A combination of the NwN information and the gland information (including gland features and gland-score) in performing automatic cancer region detection in the prostate tissue image would be of great interest.

A method to generate a confidence score when performing automatic grading of a tissue image would be very useful. By considering these scores, the pathologist only needs to manually grade the difficult images that are difficult to grade by the automatic system.

In performing cancer detection and grading, when the information contained in the H&E tissue image is not sufficiently clear, the pathologist usually obtains an additional tissue slide (from the same biopsy), stains it using the Immunohistochemistry (IHC) method (which colors the tissue components in a different way from the H&E stain), and analyzes this IHC stained tissue image. Although the two tissue images capture the same area in the biopsy (thus having similar tissue content), they may not be properly aligned with each other. Hence, an image registration method to align the H&E and IHC stained images is desired. Using this functionality, the pathologist can view the same tissue region on both H&E and IHC images to gain better information for the grading. The CAD system can also benefit from this functionality by combining the features extracted from two differently stained tissue images to improve the performance of the automatic cancer detection and grading methods.

It would be desirable to develop generalized segmentation methods to segment tissue

components (lumen, nuclei, stroma), and segment glands from the prostate tissue images of different stains. Similarly, generalized features that can be applied on images of different stains will also be useful (although specific features for each stain, e.g., the color-based features, are still necessary).

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA: A Cancer Journal for Clinicians*, vol. 63, pp. 11–30, 2013.

[2] S. Naik, S. Doyle, A. Madabhushi, J. E. Tomaszewski, and M. D. Feldman, "Automated gland segmentation and Gleason grading of prostate histology by integrating low-, high-level and domain specific information," *MIAAB workshop, Piscataway, NJ*, 2007.

[3] J. Monaco, et al., "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models," *Medical Image Analysis*, vol. 14, pp. 617–629, 2010.

[4] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern. SMC*, vol. 3, pp. 610–621, 1973.

[5] "Prostate cancer world map," `http://globocan.iarc.fr/factsheets/cancers/prostate.asp`.

[6] "The prostate in male body," `http://byhealth.com/prostate-cancer`.

[7] N. Howlader, et al., "Seer cancer statistics review, 1975-2009 (vintage 2009 populations)," Bethesda, MD, `http://seer.cancer.gov/csr/1975_2009_pops09/`, based on November 2011 SEER data submission, posted to the SEER web site, April 2012.

[8] "Digital rectal exam," `http://www.ccmurology.com/disease/benign_prostatic_hyperplasia.php`.

[9] "PSA blood test," `http://www.shealthinsurance.com/blog/2012/05/a-psa-test-may-be-doing-you-more-harm-than-good/`.

[10] "Prostate needle biopsy, ultrasound, CT scan, MRI scan," `http://ayie-disease.blogspot.com/2011/01/prostate-cancer-pictures-slideshow.html`.

[11] "Microtome machine," `http://pathology.utscavma.org/?page_id=37`.

[12] "Glass slide with tissue," `http://histology-slides-database.blogspot.com/2011/08/glass-slide-with-tissue-piece.html`.

[13] "Stained tissue slide," `http://www5.pbrc.hawaii.edu/ccrt/taras/site/histology.html`.

[14] D. Gleason, "Histologic grading and clinical staging of prostatic carcinoma," *Urologic Pathology: The Prostate, M. Tannenbaum, Ed. Philadelphia, PA: Lea and Febiger*, pp. 171–198, 1977.

[15] American Cancer Society, "Prostate cancer," 2012, `http://www.cancer.org/Cancer/ProstateCancer/`.

[16] Wikipedia, "Prostate cancer," `http://en.wikipedia.org/wiki/Prostate_cancer`.

[17] G. Steinberg, B. Carterm, T. Beaty, B. Childs, and P. Walsh, "Family history and the risk of prostate cancer," *Prostate*, vol. 17, no. 4, pp. 337–47, 1990.

[18] R. Martin, L. Vatten, D. Gunnell, and P. Romundstad, "Blood pressure and risk of prostate cancer: Cohort Norway (CONOR)," *Cancer Causes Control*, vol. 21, no. 3, pp. 463–72, 2010.

[19] V. Venkateswaran and L. H. Klotz, "Diet and prostate cancer: Mechanisms of action and implications for chemoprevention," *Nature reviews. Urology*, vol. 7, no. 8, pp. 442–453, August 2010.

[20] D. Alexander, P. Mink, C. Cushing, and B. Sceurman, "A review and meta-analysis of prospective studies of red and processed meat intake and prostate cancer," *Nutrition journal*, vol. 9, no. 50, 2010.

[21] "Computed tomography (CT) machine," `http://www.nepetimaging.com/pet-ct.html`.

[22] "Magnetic resonance imaging (MRI) machine," `http://www.webmd.com/brain/open-magnetic-resonance-imaging-mri-machine`.

[23] "Ventana Medical Systems, Inc.," `http://www.ventanadigitalpathology.com/`.

[24] W. Catalona, D. Smith, and T. Ratliff, "Measurement of prostate-specific antigen in serum as a screening test for prostate cancer," *New Engl. J. Med.*, vol. 324, no. 17, pp. 1156–1161, 1991.

[25] "Panoptiq tissue slide scanner," `http://innovative-imaging.net/panoptiq_panorama.html`.

[26] "Leica scn400 tissue slide scanner," `http://www.leicabiosystems.com/products/digital-pathology/scan/details/product/leica-scn400-2/`.

[27] "Omnyx vl120 tissue slide scanner," `http://www.omnyx.com/solutions/vl120-scanner`.

[28] "Apero tissue slide scanner," `http://www.aperio.com/services/slide/scanning`.

[29] "VS120-SL tissue slide scanner," `http://www.olympusamerica.com/seg_section/product.asp?product=1066`.

[30] "Quantifying histological data," `http://telepathologycity.com/quantifying_histological_data.htm`.

[31] F. Dick, S. Van-Lier, P. Banks, G. Frizzera, G. Witrak, R. Gibson, G. Everett, L. Schuman, P. Isacson, G. OConor, K. Cantor, W. Blattner, and A. Blair, "Use of the working formulation for non-hodgkins lymphoma in epidemiological studies: Agreement between reported diagnoses and a panel of experienced pathologists," *Journal of National Cancer Institue*, vol. 78, pp. 1137–1144, 1987.

[32] G. E. Metter, B. N. Nathwani, J. S. Burke, C. C. Winberg, R. B. Mann, M. Barcos, C. R. Kjeldsberg, C. C. Whitcomb, D. O. Dixon, T. P. Miller, and S. E. Jones, "Morphological subclassification of follicular lymphoma: Variability of diagnoses among hematopathologists, a collaborative study between the repository center and pathology panel for lymphoma clinical studies," *Journal of Clinical Oncology*, vol. 3, pp. 25–38, 1985.

[33] A. D. Ramsey, "Errors in histopathology reporting: Detection and avoidance," *Histopathology*, vol. 34, pp. 481–490, 1999.

[34] L. Teot, R. Khayat, S. Qualman, G. Reaman, and D. Parham, "The problem and promise of central pathology review: Development of a standardized procedure for the childrens oncology group," *Pediatric and Developmental Pathology*, vol. 10, pp. 199 –207, 2007.

[35] D. Gleason, "Histologic grading of prostate cancer: A perspective," *Human Pathology*, vol. 23, pp. 273–279, 1992.

[36] J. I. Epstein, "An update of the Gleason grading system," *J. Urology*, vol. 183, no. 2, pp. 433–440, 2010.

[37] K. A. Iczkowski and M. S. Lucia, "Current perspectives on Gleason grading of prostate cancer," *Current Urology Reports*, vol. 12, no. 3, pp. 216–222, 2011.

[38] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[39] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 681–685, 2001.

[40] P. Khurd, C. Bahlmann, P. Maday, A. Kamen, S. Gibbs-Strauss, E. M. Genega, and J. V. Frangioni, "Computer-aided Gleason grading of prostate cancer histopathological images using texton forests," in *International Symposium on Biomedical Imaging: From Nano to Macro*, 2010, pp. 636–639.

[41] S. K. Tai, C. Y. Li, Y. C. Wu, Y. J. Jan, and S. C. Lin, "Classification of prostatic biopsy," in *International Conference on Digital Content, Multimedia Technology and its Applications (IDC)*, August 2010, pp. 354–358.

[42] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, M. Feldman, and J. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 1284–1287.

[43] H. J. Yoon, C. C. Li, C. Christudass, R. Veltri, J. I. Epstein, and Z. Zhang, "Cardinal multiridgelet-based prostate cancer histological image classification for Gleason grading," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2011, pp. 315–320.

[44] P. Khurd, L. Grady, A. Kamen, S. Gibbs-Strauss, E. M. Genega, and J. V. Frangioni, "Network cycle features: Application to computer-aided Gleason grading of prostate cancer histopathological images," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1632–1636.

[45] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted Bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 1205–1218, 2012.

[46] K. Nguyen, A. Jain, and B. Sabata, "Prostate cancer detection: Fusion of cytological and textural features," *Journal of Pathology Informatics*, vol. 2, no. 2, pp. 2–3, 2011.

[47] V. Kumar, A. Abbas, and N. Fausto, *Robbins and Cotran Pathologic Basis of Disease*, Saunders, 2004.

[48] J. Xu, P. Monaco, and A. Madabhushi, "Markov random field driven region-based active contour model (MaRACel): Application to medical image segmentation," in *Proceedings of the 13th International Conference on Medical Image Computing and Computer-assisted Intervention: Part III*, 2010, MICCAI 2010, pp. 197–204.

[49] J. Xu, R. Sparks, A. Janowcyzk, J. E. Tomaszewski, M. D. Feldman, and A. Madabhushi, "High-throughput prostate cancer gland detection, segmentation, and classification from digitized needle core biopsies," in *Proceedings of the 2010 International Conference on Prostate Cancer Imaging: Computer-aided Diagnosis, Prognosis, and Intervention*, 2010, pp. 77–88.

[50] Y. Peng, Y. Jiang, L. Eisengart, M. Healy, F. Straus, and X. Yang, "Segmentation of prostatic glands in histology images," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 2091–2094.

[51] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: A new variational formulation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 430–436.

[52] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 48, no. 3, pp. 259–302, 1986.

[53] A. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[54] Wikipedia, "Lab color space," `http://en.wikipedia.org/wiki/Lab_color_space`.

[55] H. S. Fairman, M. H. Brill, and H. Hemmendinger, "How the CIE 1931 color-matching functions were derived from Wright-Guild data," *Color Research and Application*, vol. 22, pp. 11–23, 1997.

[56] A. Tabesh, M. Teverovskiy, and H. Pang, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1366–1378, 2007.

[57] J. A. Kiernan, *Histological and Histochemical Methods: Theory and Practice*, A Hodder Arnold Publication, 2001.

[58] R. Haralick and L. Shapiro, *Computer and Robot Vision*, vol. 2, Addison-Wesley, 1992.

[59] K. J. Khouzani and H. S. Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 697–704, June 2003.

[60] M. Pelillo, "The dynamics of nonlinear relaxation labeling processes," *J. Math. Imaging Vis.*, vol. 7, pp. 309–323, 1997.

[61] G. Wisniewski and P. Gallinari, "Relaxation labeling for selecting and exploiting efficiently non-local dependencies in sequence labeling," in *Proceedings of the 11th European conference on Principles and Practice of Knowledge Discovery in Databases*, Berlin, Heidelberg, 2007, PKDD 2007, pp. 312–323.

[62] C. Li, C. Xu, C. Gui, and M. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, 2010.

[63] J. Diamond, N. Anderson, P. Bartels, R. Montironi, and P. Hamilton, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human Pathology*, vol. 35, pp. 1121–1131, 2004.

[64] M. Teverovskiy, V. Kumar, J. Ma, A. Kotsianti, D. Verbel, A. Tabesh, H. Pang, Y. Vengrenyuk, S. Fogarasi, and O. Saidi, "Improved prediction of prostate cancer recurrence based on an automated tissue image analysis system," in *Proceedings of IEEE International Symposium On Biomedical Imaging*, 2004, pp. 257–260.

[65] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.

[66] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1150–1157.

[67] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, Washington, DC, USA, 2006, pp. 2169–2178.

[68] D. M. Kwak, B. S. Kim, O. K. Yoon, C. H. Park, J. U. Won, and K. H. Park, "Content-based ultrasound image retrieval using a coarse to fine approach," *Annals of the New York Academy of Sciences*, vol. 980, pp. 212–224, 2002.

[69] C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, and A. M. Aisen, "Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision," in *Proceedings of the 10th National Conference on Artificial Intelligence*, 1999, pp. 760–767.

[70] M. M. Rahman, B. C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Computerized Medical Imaging and Graphics*, vol. 32, no. 2, pp. 95–108, 2008.

[71] S. Beretti, A. D. Bimbo, and P. Pala, "Content-based retrieval of 3D cellular structures," in *Proceedings of the second International Conference on Multimedia and Exposition (ICME 2001), IEEE Computer Society*, Tokyo, Japan, 2001, pp. 1096–1099.

[72] A. Madabhushi, S. Doyle, M. Feldman, and J. Tomaszewski, "Malignancy diagnosis using content-based image retrieval of tissue histopathology," *US Patent 2010/0098306*, April 2010.

[73] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[74] B. Parvin, Q. Yang, J. Han, H. Chang, B. Rydberg, and M. Barcellos-hoff, "Iterative voting for inference of structural saliency and characterization of subcellular events," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 615–623, 2007.

[75] R. Balakrishnan, G. Sethuraman, and R.J. Wilson, *Graph Theory And Its Applications*, Narosa Publishing House, 2004.

[76] C.T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," *IEEE Transactions on Computers*, vol. C-20, no. 1, pp. 68–86, 1971.

[77] "Ellipse fitting implementation," `www.mathworks.com/matlabcentral/fileexchange/3215-fitellipse`.

[78] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[79] M. Oussalah, "Content based image retrieval: Review of state of art and future directions," in *First Workshops on Image Processing Theory, Tools and Applications*, 2008, pp. 1–10.

[80] T. Kato, "Database architecture for content-based image retrieval," *Image Storage and Retrieval Systems, Proc SPIE 1662*, pp. 112–123, 1992.

[81] M. Flickner, et al., "Query by image and video content: the QBIC system," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[82] M. Gupta, "The Virage image search engine: An open framework for image management," *Storage and Retrieval for Image and Video Databases IV, Proc SPIE 2670*, pp. 76–87, 1996.

[83] Y. Rui and T. Huang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 39–62, 1999.

[84] R. Datta, D. Joshi, J. Li, and J. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 34–94, 2008.

[85] A. Jain, R. Jin, and J. Lee, "Tattoo image matching and retrieval," *IEEE Computer*, vol. 45, no. 5, pp. 93–96, 2012.

[86] M. Bolle, H. Connell, N. Haas, R. Mohan, and G. Taubin, "VeggieVision: A produce recognition system," in *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96)*, 1996, pp. 244–251.

[87] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications - clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, pp. 1–23, 2004.

[88] W. Cai, D. Feng, and R. Fulton, "Content-based retrieval of dynamic PET functional images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 4, no. 2, pp. 152 –158, 2000.

[89] F. Schnorrenberg, C. S. Pattichis, C. N. Schizas, and K. Kyriacou, "Content-based retrieval of breast cancer biopsy slides," *Technol. Health Care*, vol. 8, no. 5, pp. 291–297, 2000.

[90] U. Sinha and H. Kangarloo, "Principal component analysis for content-based image retrieval," *Radio Graphics*, vol. 22, no. 5, pp. 1271–1289, 2002.

[91] J. C. Caicedo, F. A. Gonzalez, and E. Romero, "Content-based histopathology image retrieval using a kernel-based semantic annotation framework," *Journal of Biomedical Informatics*, vol. 44, no. 4, pp. 519 – 528, 2011.

[92] D. J. Rogers and T. T. Tanimoto, "A computer program for classifying plants," *Science*, vol. 132, pp. 1115–1118, 1960.

[93] Y. Huang, J. Zhang, Y. Zhao, and D. Ma, "Medical image retrieval with query-dependent feature fusion based on one-class SVM," in *2010 IEEE 13th International Conference on Computational Science and Engineering (CSE)*, 2010, pp. 176–183.

[94] H. Tamura, S. Mori and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[95] Y. Song, W. Cai, and D. Feng, "Hierarchical spatial matching for medical image retrieval," in *Proceedings of the 2011 International ACM Workshop on Medical Multimedia Analysis and Retrieval*, 2011, pp. 1–6.

[96] R. Sparks and A. Madabhushi, "Content-based image retrieval utilizing explicit shape descriptors: Applications to breast MRI and prostate histopathology," in *SPIE Medical Imaging*, 2011, pp. 7962–79621I.

[97] M. Ortega, Y. Rui, K. Chakrabarti, K. Porkaew, S. Mehrotra, and T. S. Huang, "Supporting ranked Boolean similarity queries in MARS," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 6, pp. 905 –925, 1998.

[98] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: Integrated region matching for image retrieval," in *Proc. of ACM Multimedia*, 2000, pp. 147–156.

[99] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, vol. 24, pp. 1167–1186, 1991.

[100] L. H. Tang, R. Hanka, R. Lan, and H. Ip, "Automatic semantic labelling of medical images for content-based retrieval," in *Proceedings of the International Conference on Artificial Intelligence, Expert Systems and Applications*, 1998, pp. 77–82.

[101] L. H. Tang, R. Hanka, H. Ip, and R. Lam, "Extraction of semantic features of histological images for content-based retrieval of images," in *Proceedings of the IEEE Symposium on Computer-Based Medical Systems (CBMS 2000)*, Houston, TX, USA, 2000.

[102] H. Müller, A. Rosset, J. Vallée, and A. Geissbuhler, "Integrating content-based visual access methods into a medical case database," in *Proceedings of the Medical Informatics Europe Conference (MIE 2003)*, St. Malo, France, 2003, vol. 95, pp. 480–485.

[103] D. M. Squire, W. Muller, H. Muller, and J. Raki, "Content-based query of image databases, inspirations from text retrieval: Inverted files, frequency-based weights and relevance feedback," in *The 10th Scandinavian Conference on Image Analysis (SCIA'99)*, 1999, pp. 143–149.

[104] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[105] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Medical Image Analysis*, vol. 14, pp. 227–241, 2010.

[106] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *Computer Vision and Image Understanding*, vol. 84, no. 1, pp. 25–43, 2001.

[107] B. Helpap and J. Kollermann, "Combined histoarchitectural and cytological biopsy grading improves grading accuracy in low-grade prostate cancer," *International Journal of Urology*, vol. 19, no. 2, pp. 126–133, 2012.

[108] Y. Gong, N. Caraway, J. Stewart, and G. Staerkel, "Metastatic ductal adenocarcinoma of the prostate: Cytologic features and clinical findings," *J. clin. Path.*, vol. 126, pp. 302–309, 2006.

[109] B. Helpap, "Cell kinetic and cytological grading of prostatic carcinoma," *Virchows Archiv*, vol. 393, pp. 205–214, 1981.

[110] R. W. Veltri, S. Isharwal, M. C. Miller, J. I. Epstein, L. A. Mangold, E. Humphreys, and A. W. Partin, "Long-term assessment of prostate cancer progression free survival: Evaluation of pathological parameters, nuclear shape and molecular biomarkers of pathogenesis," *Prostate*, vol. 68, no. 16, pp. 1806–1815, 2008.

[111] M. Mason, "Cytology of the prostate," *J. Clin. Path.*, vol. 17, no. 6, pp. 581–590, 1964.

[112] S. Diaconescu, D. Diaconescu, and S. Toma, "Nucleolar morphometry in prostate cancer," *Bulletin of the Transilvania University of Brasov*, vol. 3, 2010.

[113] "Matlab function to create a grayscale image," `http://www.mathworks.com/help/toolbox/images/ref/rgb2gray.html`.

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