Stony Brook University

Cancer Diagnosis and Prognosis with Histopathology Image Analysis and Pattern Recognition

A Dissertation presented

by

Naiyun Zhou

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Biomedical Engineering

Stony Brook University

December 2017

Stony Brook University

The Graduate School

Naiyun Zhou

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Chuan Huang - Dissertation Advisor
Assistant Professor, Radiology and Psychiatry
Program Faculty, Biomedical Engineering

Paul Vaska - Chairperson of Defense
Professor, Biomedical Engineering

Shu Jia - Committee Member
Assistant Professor, Biomedical Engineering

Wei Zhu - Committee Member
Professor, Applied Mathematics and Statistics

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Cancer Diagnosis and Prognosis with Histopathology Image Analysis and Pattern Recognition

by

Naiyun Zhou

Doctor of Philosophy

in

Biomedical Engineering

Stony Brook University

2017

The dissertation focuses on the histopathology image analysis for cancer diagnosis and prognosis, based on the innovation of histopathological machine vision techniques. Histopathology images are regarded as the reference standard to identify diseases, and especially as a gold standard on cancer diagnosis. With the recent advance of the electronic scanners, digitized whole slide images (WSI) make it possible to analyze cancer tissues in high resolution and large-scale manner. At the same time, computer aided diagnosis (CAD) algorithms are being developed to detect cancer automatically, both in radiological and pathological field. However, those CAD algorithms are based on object segmentation and handcrafted features, which are not fully automatic. I developed innovative methods and frameworks to assist cancer diagnosis and prognosis automatically, without sophisticated feature extraction. The major projects are intermediate prostate cancer grading, cell nuclei segmentation using deep learning and nuclei segmentation evaluation through image synthesis. My research novelties include multi-resolution histopathology image analysis, fully automatic gland cancerous degree classification, and nuclei segmentation and synthesis using deep learning.

Contents

List of Figures

List of Tables

Acknowledgements

The six and a half year of my Ph.D. career at Stony Brook is a lucky and hard experience in my life. First of all, I express deep thanks to my research advisor, Professor Yi Gao for giving me precious research opportunity and academically professional guidance. He led me to the field of medical image computing, taught me cutting-edge technology and gave me so much research freedom that I will be grateful forever. I would also like to thank my co-advisor (who is responsible for administrative matter), Professor Chuan Huang, who encouraged and supported me when I almost lost my hope to continue my study. I am thankful to my committee members, Professor Paul Vaska, Professor Shu Jia and Professor Wei Zhu for their valuable suggestion and inspiration. In addition, I would like to thank Professor Klaus Mueller for his encouragement that I should pursue my dream.

I am grateful towards my wonderful collaborators, fellow graduate students and great friends I have made at this journey. I would like to thank Cheng Chang, Anzhou Zhang, Si Wen, Le Hou, Maozheng Zhao, Zhiyuan Zhang, Bing Wang, Shun Yao, Heyi Li, Yifan Hu, Hao Zhang and Hao Han, for their enlightening discussions and collaborations on medical image analysis and artificial intelligence. I am indebted to Youquan Chong, Jian Xu, Cong Xie and Wen Zhong, for their altruistic guidance and help in technical respect. I would also like to appreciate the help from Yulin Huang, Xuyi Wang, Kun Ma, Yiyao Tian, Lizhou Nie, Jiang You, Wei Chen, Qiujia Zhang, Ye Chen, Xin Wei and Xin Yue in my campus life.

Last but not Least, I am deeply grateful for my parents and my boyfriend for their unconditional love to me. You always make me feel warm along the way of my Ph.D. pursuit.

Publications

1. ***Naiyun Zhou***, *Xiaxia Yu, Tianhao Zhao, Si Wen, Fusheng Wang, Wei Zhu, Tahsin Kurc, Allen Tannenbaum, Joel Saltz, Yi Gao. "Evaluation of Nucleus Segmentation in Digital Pathology Images through Large Scale Image Synthesis."* SPIE Medical Imaging Conference. Orlando, Florida, February 2017. **Oral presentation**.

2. ***Naiyun Zhou***, *Yi Gao. "Optimized Color Decomposition of Localized Whole Slide Images and Convolutional Neural Network for Intermediate Prostate Cancer Classification."* SPIE Medical Imaging Conference. Orlando, Florida, February 2017. **Best Student Paper Award Finalist**.

3. ***Naiyun Zhou***, *Joel Saltz, Klaus Muller.* Maps of Human Disease: A Web-Based Framework for the Visualization of Human Disease Comorbidity and Clinical Profile Overlay. *The First International Workshop on Data Management and Analytics for Medicine and Healthcare, in conjunction with VLDB 2015.*

4. *Cheng Chang, Chuan Huang,* ***Naiyun Zhou****, Lawrence Ver Hoef, Yi Gao. "The Bumps Under the Hippocampus."* Human Brain Mapping*, DOI: 10.1002/hbm.23856, 2017.*

5. *Shun Yao, Cheng Chang, Wei Xu,* ***Naiyun Zhou****.* NNLSF: A Fast and Informative Fitting Method for XANES Chemical Mapping Analysis. *2015 IEEE International Symposium on Biomedical Imaging*

6. *Pengyue Zhang, Huixian Yu,* ***Naiyun Zhou****.* Early Exercise Improves Cerebral Blood Flow Through Increased Angiogenesis in Experimental Stroke Rat Model. *Journal of NeuroEngineering and Rehabilitation*

Chapter 1

Introduction

1.1 Background and Significance

Cancer is the second leading cause of death. It is a class of diseases caused by the uncontrolled growth and expansion of abnormal cells. In 2015, over 1,600,000 new cancer cases were expected in United States and over 500,000 Americans were expected to die of cancer [1]. There is no doubt that cancer is the most important public health problem in the United States.

Since the mid-1990s, the overall cancer mortality rates have dropped steadily, which is attributable to improvement in diagnosis and treatment [2]. The computer-aided analytical approaches to radiological and histopathological images have developed a lot over the past decade. On one hand, CAD relieves the workload on pathologists by removing the most obvious benign areas. On the other hand, quantitative tissue analysis alleviates intra- and inter-observation variations (subjective opinions) among pathologists, yielding a more objective basis for treatment.

However, there remain difficulties and challenges in CAD, especially in the field of histopathology image analysis. For example, tissue structures in cancerous conditions provide a wealth of information, which makes object segmentation extremely difficult. In addition, there is a large space for the research of disease-specific characteristics to improve spatial analysis of histopathology imagery [3]. Last but not least, almost research on useful features for machine learning are handcrafted, which is inflexible and not

specific. There is no general framework for histopathological image classification due to too many parameters and method options in the automated detection system.

Recently, deep learning architectures have achieved great success in computer vision and pattern recognition field. In contrast to human feature engineering, convolutional neural networks (CNNs), as fully data driven approaches, have shown outstanding performance in various tough computer vision tasks [4, 5, 6]. CNN is computationally complex and needs large amount of training set. But when we use GPUs for CNNs training, we can get a lot speedup. Moreover, the digitized WSIs are often several gigabytes in size. Once WSIs are labeled, the amount of image patches extracted from a single WSI is on the order of thousands. Therefore, CNN is suitable for automated classification and feature recognition for cancer histopathological images.

In general, cancer histopathology image analysis is of high importance in diagnosis, prognosis and treatment. Computer aided diagnosis (CAD) algorithms are being developed to detect cancer automatically, both in radiological and histopathological field. However, those CAD algorithms are based on object segmentation and handcrafted features, which are not fully automatic. Given abundant histopathological images in TCGA database, state-of-art CNN architectures, and powerful GPU computation resources, I developed appropriate and efficient histopathological image analysis frameworks to solve cancer diagnosis and prognosis issues, thus alleviating human labor, and intra-,inter-subjective opinions among pathologists.

1.2 Outline

The dissertation focuses on the histopathology image analysis for cancer diagnosis and prognosis, based on the innovation of histopathological machine vision techniques. I developed innovative methods and frameworks to assist cancer diagnosis and prognosis automatically, without sophisticated feature extraction. The major projects are automatic cancer grading, histopathological segmentation using deep learning and evaluation of nuclei segmentation through image synthesis. My research novelty includes multi-resolution

histopathology image analysis, gland cancerous degree classification, cell nuclei segmentation for different types of cancer, and nuclei image synthesis with ground truth mask.

The outline of the proposal is as follows: Chapter 2, Image Analysis Techniques and Models. Histopathology Image Analysis introduces the basic methods of the image preprocessing, automated object detection and classification of histopathology images. Pattern Recognition Techniques part provides the fundamental algorithms in machine learning and deep learning that are useful in my research, and the corresponding optimization algorithms. Chapter 3, Intermediate Prostate Cancer Classification, is a fully automatic approach to grade intermediate prostate malignancy. The content is the second paper in the Publication List. Chapter 4, Nuclei Segmentation for Various Types of Cancer, is a challenge solution for nuclei segmentation in four types of cancer histopathology images. Our deep learning method outperform the challenge-winning result. The content is the first paper in the Manuscript List. Chapter 5, Evaluation of Nuclei Segmentation via Image Synthesis, which is an idea of computationally generating ground truth for nuclei segmentation in histopathology images, to solve the manual annotation issue. The content is the first paper in the Publication List. Chapter 6, Evaluation of Nuclei Segmentation via Image Synthesis, is a deep learning approach based image synthesis for nuclei segmentation evaluation. The content is the third paper in the Manuscript List. Appendix A, Clinical Information Visualization, is a side project on clinical data visualization algorithms. This work was published before my histopathology image analysis research work. It is a web-based framework for the visualization of human disease comorbidity and clinical profile overlay. The clinical data is from general medicine. Once we obtain cancer clinical data, we could leverage this framework to carry our further visualization or clinical data mining research. The content is the third paper in the Publication List. Appendix B, Digital Pathology Image Viewing Module, is an implementation of a large-scale pathology image viewing and editing module in 3D Slicer, which is utilized for image acquisition step in Chapter 3-6.

Chapter 2

Image Analysis and Pattern Recognition

2.1 Histopathology Image Analysis

This section will give an introduction to the fundamental techniques in histopathology image analysis. Color decomposition is the image preprocessing method to extract target tissue composition. Object detection and segmentation refer in particular to nuclear and gland segmentation because certain characteristics of them are hallmarks of cancerous conditions. Feature extraction and classification section consists of multi-scale feature selection, dimensionality reduction and discriminative classifiers.

2.1.1 Color Decomposition

Histopathology refers to the study of the microscopic anatomical changes in diseased tissue. By using the microscopic examination of a biopsy or surgical specimen, pathologists process the diseased tissue and fixed them onto a glass slides. The most important step of the process is staining, which is employed to give both contrasts to the tissue as well as revealing cellular components. Hematoxylin and eosin (H&E stain) is the most commonly used light microscopical stain for simple morphological diagnosis in histopathology. Hematoxylin, a basic dye, stains nuclei blue due to an affinity to nucleic acids in the cell nucleus. Eosin, an acidic dye, stains the cytoplasm pink. Figure 2.1 is an image of human prostate tissue H&E stain. We can tell

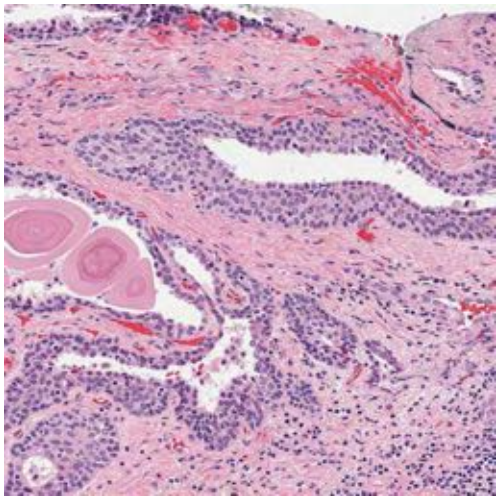


Figure 2.1: A human prostate tissue H&E stain.

that, hematoxylin stains cell nuclei blue, while Eosin stains cytoplasm and connective tissue pink. Because of the bright future of H&E stain [7], most of the histopathology images used in my thesis are digitized H&E stained sections.

In order to diagnose disease with differential staining, we would indicate the amount of the substances, eg., cell nuclei , cytoplasm and other cell organelles, or structures to which the stain specifically attaches. In histopathology staining, there is partial overlap in the absorption spectra of different dyes, so we cannot use single wavelength to quantify each stain [8]. RGB information in each pixel is determined by the stain concentrations according to the Lambert-Beer law [9]. Light transmitting though matter is attenuated by a ratio, defined as the absorption coefficient A_c. Assuming the medium is homogeneous, we can compute the attenuation of the light by Lambert-Beer′s law:

$$I_c = I_{0,c} \cdot exp(-A_c) \tag{2.1}$$

where $I_{0,c}$ is the intensity of the light entering the medium, I_c is the intensity of the light detected after passing the medium.

Although the intensities I_R, I_G, I_B are in a nonlinear relationship with the concentration of stain, the optical density for each channel is linear with the

concentration of the stain, which can be used for quantification of multiple stains. The optical density (OD) in each of the channels (RGB) can be defined as:

$$OD_c = -log(I_c/I_{0,c}) = A_c \tag{2.2}$$

where c indicates the detection channel.

Dr. Arnout Ruifrok proposed a quantification of histochemical staining by color deconvolution in 2000 [10]. In his paper, each pure stain (hematoxylin, eosin and diaminobenzidine) is characterized by a specific OD for the light through each of the three RGB channels. Here we represent the OD vector as a 1 by 3 vector, eg., M_h for hematoxylin, describing each stain in the OD-converted RGB color space.

$$M_h = [h_R, h_G, h_B],$$

where h_R, h_G and h_B stand for hematoxylin-specific values for the OD in each of the three RGB channels. Here, $h_R = 0.18$, $h_G = 0.20$ and $h_B = 0.08$. Obviously, we can represent the OD for all three channels by a matrix, M, consisting three OD vectors:

$$M = \begin{bmatrix} M_h \\ M_e \\ M_d \end{bmatrix},$$

where $M_e = (0.01, 0.13, 0.01)$, $M_h = (0.10, 0.21, 0.29)$. In order to perform separation of the stains by orthonormal transformation of the RGB information, we have to normalize the OD matrix M, to achieve correct balancing of the absorption factor for each stain, resulting in a normalized OD matrix, M_N:

$$M_N = \begin{bmatrix} 0.65 & 0.70 & 0.29 \\ 0.07 & 0.99 & 0.11 \\ 0.27 & 0.57 & 0.78 \end{bmatrix},$$

If we represent the image by O and the amounts of three stains on each pixel by C, we can get the relationship:

$$O = C \cdot M_N, \tag{2.3}$$
$$where\ C = [C_h, C_e, C_d].$$

Therefore, we can obtain the vector C, which is the quantification result, with the inverse of the OD matrix:

$$C = M_N^{-1} \cdot O, \tag{2.4}$$

$$where\ M_N^{-1} = \begin{bmatrix} 1.88 & -0.07 & -0.60 \\ -1.02 & 1.13 & -0.48 \\ -0.55 & -0.13 & 1.57 \end{bmatrix}.$$

After Dr.Ruifrok's idea, some blind color determination methods were developed which do not require reference color input. Dr. Rabinovich presented the first automated system for performing color decomposition by unsupervised learning [11]. This work investigated the use of two techniques: NMF (Non-negative Matrix Factorization) and ICA(Independent Component Analysis) by comparing the performance with ground truth data.

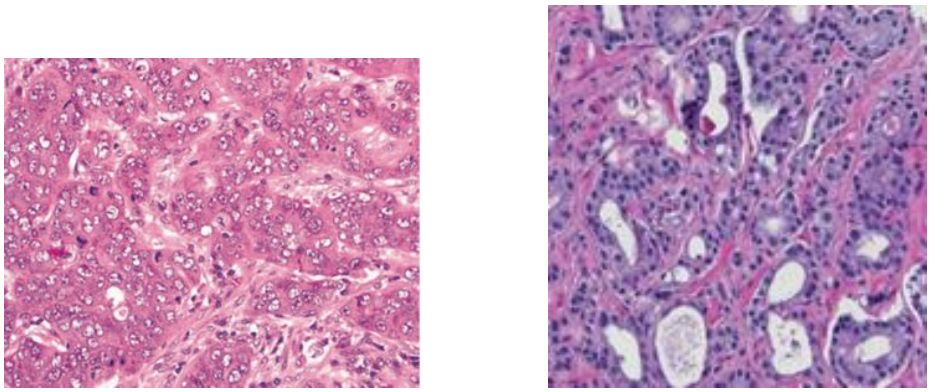
More recently, Dr. Gavrilovic developed a blind color decomposition method using statistical techniques for noise modeling of the CCD array and biochemical noise [12]. They demonstrated the method qualitatively and quantitatively showing more accurate decompositions than Dr. Rabinovich's NMF and ICA.

2.1.2 Object Detection and Segmentation

The prerequisite to disease identification and classification in histopathology images is the segmentation of objects of interest from the background. A large amount of works have been conducted on breast cell nuclear segmentation and prostate gland segmentation because of their importance in diagnosis and grading of cancer.

Figure 2.2 shows two example tissue of H&E stains from a common breast cancer and a grade 4 prostate cancer. Breast cancer and prostate cancer are the most prevalent form of cancer among women and men respectively. The aggressive potential of the breast tumor depends on the cell nuclear features and cell mitotic activity. For prostate cancer grading, the Gleason Pattern 1-5 depends on how well-formed the glands are. Therefore, in breast cancer and prostate cancer CAD systems, the segmentation of breast cells and the

prostate glands is a significant processing step. Algorithms like Bayesian classifier, level set and active contour are most common ones.



(a) (b)

Figure 2.2: (a) Invasive Ductal Carcinoma of Breast [13]. (b) Grade 4 Prostate Cancer [14].

Bayesian Classifier

For both application, a Bayesian classifier was proposed based on pixel values to segment nuclear and glandular structures [15]. It is a low-level information extraction methodology which can be utilized as a preprocessing step for high-level methodologies described later. In a digital image $S = f(s)$, where s denotes a 2 dimensional grid of pixels and f is the function that gives values (RGB channels) to image pixels. A cell comprises two main structures: cytoplasm and nuclei, which are denoted by a set $T = \{C, N\}$. In the training set T_v, the pixel values $f(s)$ are known to obtain the probability density functions $p(f(s)|v)$, where v stands for the class C or N. We employ the theorem of Bayesian Classifier to get the posterior conditional probability that s belongs to v,

$$P(v|f(s)) = \frac{P(v)p(f(s)|v)}{\sum_{w\in\{C,N\}} P(w)p(f(s)|w)} \tag{2.5}$$

where $p(f(s)|v)$ is the priori conditional probability obtained in training set, $v \in \{C, N\}$. $P(v)$ is the prior probability of occurrence for each class, like

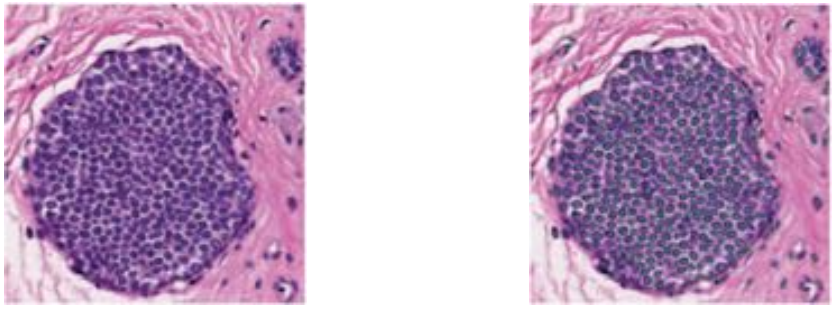
cytoplasm or nuclei. The results are the pixel-wise likelihood scenes, which are the probability of pixel s belonging to class v. We can set an empirically predefined threshold to get the binary image for nuclei. In the prostate gland case, class v will consist of lumen, cytoplasm and nuclei.

Level Set

Level set based deformable methods have been widely used for medical image object segmentation. In mathematics, a level set of a real valued function f of n real variables is a set of the form: $L_c(f) = \{(x_1, ..., x_n)|f(x_1, ..., x_n) = c\}$. In 2 dimensions, the level set method represents a closed curve P (like the boundary of nuclear or gland) by using an auxiliary function ϕ that is the level set function. Mathematically, P can be expressed as $P = \{(x, y)|\phi(t, x, y) = 0\}$ If the curve P moves in the normal direction with a speed v, then the level set function ϕ satisfies the equation:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0 \tag{2.6}$$

where the function F defines the speed of the movement. For a histopathology image segmentation application [15], the initial contour $\phi_0 = \phi(0, x, y)$ is defined by the Bayesian classifier. Figure 2.3 is the original image and the segmentation result.



(a) (b)

Figure 2.3: (a) Original Image (b) Segmentation Result [15].

Active Contour

The active contour in this section refers to the snake class, also called explicit or parametric active contour. The snake is an energy-minimizing spline constrained by image forces, external constraint forces and internal spline forces to detect edges or lines [16]:

$$E^*_{snake} = \int_0^1 E_{snake}(v(s))ds = \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))ds \tag{2.7}$$

where $v(s) = (x(s), y(s))$ is the position in an image, E_{int} represents the internal energy of the spline due to bending, E_{image} changes due to the image forces and E_{con} changes due to the external constraint forces. The minimization procedure is an iterative technique using sparse matrix methods. In terms of the internal energy, each iteration effectively takes implicit Euler steps, while for the image and external constraint energy, explicit Euler steps were taken. By deforming a curve of sharp image intensity variations, the segmentation of any object can be determined.

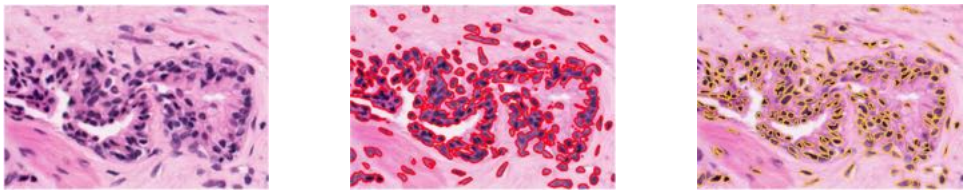
In 2012, Dr. Madabhushi published a novel boundary and region-based active contour model to resolve boundaries of overlapped objects [17]. It incorporates shape priors with the form of level set algorithm that is initialized automatically based on watershed. The energy functional of the novel active contour:

$$F = \beta_1 F_{shape}(C) + \beta_2 F_{boundary}(\phi, \psi) + F_{region}(\phi, u_{in}, u_{out}) \tag{2.8}$$

where ϕ is the level set function, ψ is the shape function of the object, u_{in} and u_{out} are the partitioned foreground and background regions, and β_1 and β_2 are the weights balancing the contributions of the boundary, shape and region terms. The algorithm is able to resolve up to 91% of overlapping nuclear and glandular structures on digitized histopathology images of 14 breast and 100 prostate biopsy specimens. Figure 2.4 reflects the superiority of nuclear segmentation of the model.

2.1.3 Feature Extraction and Classification

Research on image feature extraction and selection have been developed in analysis of histopathology imagery. The image feature itself and the selection



(a) (b) (c)

Figure 2.4: (a) Original histological image. (b) Segmentation results from geodesic active contour model. (c) Segmentation results from the novel active contour model. [17]

algorithm are both crucial for classification (disease grading and diagnosis). In this section, histopathology image features will be introduced in object level and spatially related level at first. After feature selection, dimensionality reduction techniques will be introduced. At last, disease discrimination and grade quantification are implemented by classification frameworks.

Object and Spatial-Arrangement Features

In a CAD system, image feature extraction and selection follow the segmentation, which is also crucial for pattern recognition. For the regions of interest, such as nuclear and gland, object-level features can to classified into four groups. They are listed in Figure 2.5: 1) Size and Shape, 2) Radiometric and Densitometric, 3) Texture, and 4) Chromatin-specific. These low-level metrics that depends on the quality of segmentation can be combined with statistics to generate very high dimensional feature vectors for a single object.

Different from traditional object features, topological or graph-based features represent spatial structure. Voronoi diagrams, Delaunay triangulation and minimum spanning trees are most popular graph data structures extracted from histopathology image for classification and prediction. A very complete summary of all the spatial features published in the literature is listed in Figure 2.6. The graph data structure constructed from histopathology tissue objects are highly related to pathology states. For example, in [18], Voronoi diagrams, Delaunay triangulation and minimum spanning trees were constructed from the centers of each nuclei. They were used to quantify

Category	Features
Size and Shape	Area
	Elliptical Features: Major and minor axis length, eccentricity, orientation, elliptical deviation
	Convex Hull Features: Convex area, convex deficiency, solidity
	Filled Image Features: Filled area, Euler number
	Bounding Box Features: Extent, aspect ratio
	Boundary Features: Perimeter, radii, perimeter Fourier energies, perimeter curvature, bending energy, perimeter fractal dimension
	Other Shape Features: Equivalent diameter, sphericity, compactness, inertia shape
	Center of Mass
	Reflection Symmetry
Radiometric and Densitometric	Image Bands, Intensity
	Optical density, integrated optical density, and mean optical
	Hue
Texture	Co-occurrence Matrix Features: Inertial, energy, entropy, Homogeneity, maximum probability, cluster shade, cluster
	Fractal Dimension
	Run-length Features: Short runs emphasis, long runs emphasis, gray-level non-uniformity, run-length non-uniformity, runs percentage, low gray-level runs emphasis, high gray-level runs
	Wavelets Features: Energies of detail and low resolution images
	Entropy
Chromatin-Specific	Area, integrated optical density, mean optical density, number of regions, compactness, distance, center of mass

Figure 2.5: Summary of object-level features used in histopathology image analysis[3].

the arrangement of nuclei and glandular structures with a support vector machine to distinguish between all four types of tissue for cancer diagnosis.

Dimensionality Reduction

With large amount of features extracted from large-scale histopathology images, vast quantity of data to be processed can be prohibitive for feasible classification. Even using current high performance computing machines, dimensionality reduction techniques are necessary to select features that optimize classification performance. Most commonly used methods are principal component analysis (PCA), Linear Discriminant Analysis (LDA), Multidimensional Scaling (MDS) and Independent Component Analysis (ICA).

Graph Structure	Features
Voronoi Tesselation	Number of nodes, number of edges, cyclomatic number, Number of triangles, number of k-walks, spectral radius, Eigenexponent, Randic index, area, roundness factor, area disorder, roundness factor homogeneity
Delaunay Triangulation	Number of nodes, edge length, degree, number of edges, Cyclomatic number, number of triangles, number of k-walks, Spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension
Minimum Spanning Tree	Number of nodes, edge length, degree number of neighbors, Wiener index, eccentricity, Randic index, Blaban index, fractal dimension
O'Callaghan Neighborhood Graph	Number of nodes, number of edges, cyclomatic number, number of neighbors, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, fractal dimension
Connected Graph	Number of nodes, edge length, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension
Relative Neighbor Graph	Number of nodes, number of edges, cyclomatic number, number of neighbors, number of triangles, number of k-walks, spectral radius, eigenexponent, Randic index, fractal dimension
K-NN Graph	Number of nodes, edge length, degree, number of triangles, number of k-walks, spectral radius, eigenexponent, Wiener index, eccentricity, Randic index, fractal dimension

Figure 2.6: Summary of spatial-arrangement features used in histopathology image analysis[3].

Classification Methods

Support Vector Machines (SVM) [18] and AdaBoost algorithm [19] have been applied most frequently to histopathology image classification for cancer grading. More interesting framework has also been proposed based on Gabor filter texture features [20], with 88% accuracy in Gleason score classification for prostate cancer. A shape-based template matching algorithm [15] was applied in automated nuclei segmentation for breast cancer. It shown a over 80% accuracy in breast cancer grade classification in which the choice of template is inspired by the nuclei features.

2.2 Pattern Recognition Models and Optimization

In histopathology image analysis, numerous machine learning and pattern recognition techniques have been applied in object detection and disease classification. In this section, I have selected several efficient models for my research of histopathology digital image analysis. Clustering algorithms

provide the ways to assign data points to a number of components parametrically or non-parametrically. The clustering of pixels might be a step in an integrated framework while plays a role in detailed annotation of cancer cells. Markov Random Fields (MRF) are undirected graphical models which are sets of random variables representing dependencies among nodes. The MRF inference is to find the most appropriate labeling which is a discrete set of labels for image pixels such that the corresponding objective function is minimized. MRFs can deal with various fundamental problems in medical image analysis such as segmentation, registration and shape alignment, etc. Convolutional Neural Networks (CNN), in contrast to human feature engineering, are fully data driven approaches. As a deep learning model, CNN has a large advantage in pattern recognition for large-scale histopathology image analysis.

2.2.1 Clustering

In this section, I will introduce two common clustering techniques: K-means algorithm and Gaussian mixture model.

K-means Clustering

Suppose we have an image data set $\{x_1, ..., x_N\}$ consisting of N pixels of 3-dimensional (RGB) variable x. Our task is to partition the pixels into K clusters by Euclidean distance between RGB variables. Then we can define the K centers by a vector $C_K = \{c_1, ..., c_K\}$. The problem can be described as finding an assignment of image pixels to clusters (a center c_k in C_K) such that the sum of the squares of the distances of each image pixel to its closest center C_K is a minimum. For each image pixel x_n, we define an indicator variable $i_{nk} = 0$ *or* 1, describing which one of the K clusters the image pixel n is assigned to.

$$F = \sum_{n=1}^{N} \sum_{k=1}^{K} i_{nk} ||x_n - c_k||^2 \tag{2.9}$$

We can find values for the i_{nk} and C_K by an iterative procedure to minimize F. In each iteration, there are two steps: 1) assignment step, 2) update step. Initially, we choose some random values for the C_K. In the assignment step, each pixel is assigned to the center whose Euclidean distance is the least. In

the update step, new means of the pixels assigned to the same center are set to be the new centers.

Gaussian Mixture Model

The Gaussian mixture model of a variable is defined as a liner superposition of Gaussian components,

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{2.10}$$

where the mixing coefficients π_k is the probability of cluster k that the variable is assigned to,

$$p(z_k = 1) = \pi_k \in [0, 1]$$

where z is a random binary variable, 1-of-K representation, in which a particular element z_k is equal to 1 and all other elements are equal to 0.

Therefore, the probability of x given the z is a product of Gaussian mixtures,

$$p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k}.$$

Since we have conditional probability theorem: $p(x, z) = p(z)p(x|z)$, we have the model as described in (2.10):

$$p(x) = \sum_{z} p(z)p(x|z) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k).$$

The model has a discrete latent variable z, which means for every observation x_n, there is a corresponding coefficients z_n. In order to find the maximum likelihood, we need to calculate the posterior probability of z given x,

$$\beta(z_k) = p(z_k = 1|x) = \frac{p(x, z_1 = 1)}{p(x)} = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x|z_j = 1)}$$

So,

$$\beta(z_k) = \frac{\pi_k N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} p_j N(x|\mu_j, \Sigma_j)}. \tag{2.11}$$

By using Expectation Maximization algorithm, we obtain ($N_k = \sum_{n=1}^{N} \beta(z_{nk})$):

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \beta(z_{nk}) x_n,\ \Sigma = \frac{1}{N_k} \sum_{n=1}^{N} \beta(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T,\ \pi_k = \frac{N_k}{N}. \tag{2.12}$$

If we define the data set of observations is $\{x_1, ..., x_n\}$, and assume that the image pixels are independently from the distribution, the log likelihood function:

$$ln\ p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} ln\{\sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k)\}. \tag{2.13}$$

The EM for the case of the GMM is to first choose initialization for the μ, Σ and π. Then alternating between 2.11 and 2.12, which are the Expectation step and the Maximization step respectively. After each iteration, we should evaluate 2.13 to guarantee the convergence of the log likelihood.

2.2.2 Markov Random Fields

A Markov Random Field (MRF) is an undirected graphical model that specifies both a factorization and a set of conditional independence relations. In undirected graphs, it is convenient to discuss conditional independence properties because there is no link between any pair of nodes. In image analysis, each node associates a pixel of an image. A MRF is formally defined by a graph $G = (V, E)$:

- $V = \{1, 2, ..., N\}$, a set of nodes, each of which is associated with a random variable w_n, $n = 1, 2, ..., N$.
- A set of neighbors of node i, $S = \{S_1, S_2, ..., S_N\}$, any element S_i which have adjacent S_j satisfies $(i, j) \in E$.
- $p(w_i|\{w_j\}_{j \in V \setminus i}) = p(w_n|\{w_n\}_{j \in S_i})$, where S_i is the Markov blanket of node i. This property is called the Markov property.

The MRF model can be described as the joint probability of the variables ($\{w_n\}$) as the product of potential functions defined on maximal cliques of G:

$$p(w) = \frac{1}{Z} exp(-E(w, \theta)),\ where\ E(w, \theta) = \sum_{m \in M} \Psi_m(\bar{w}_m, \theta_m)$$

where M is the set of maximal cliques of the graph; Ψ_m is the clique potential, $m \in M$, a potential function returns a non-negative value parameterized by θ_m; Z is the partition function that is a normalizing constant ensuring the result is a valid probability distribution, $Z = \sum_{w_1 \ldots w_N} \prod_{m \in M} exp(-\Psi_m(\bar{w}_m, \theta_m))$.

The MRF can be applied in many vision tasks such as de-noising, texture synthesis, segmentation, etc. In histopathology image analysis, MRF has been developed to perform the prostate cancer segmentation [21], detection of prostate cancer [22] and region-based active contour model for cancer lesion segmentation [23].

2.2.3 Convolutional Neural Network

Motivation

Convolutional Neural Network (CNN) has achieved tremendous success in computer vision applications. The name "convolutional neural network" indicates that the network employs a mathematical operation called convolution. It is a specialized kind of linear operation and it is the core building block of CNNs. Generally speaking, convolutional neural networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their operation layers. If we now have an input two-dimensional image I, and a two-dimensional kernel K, the discrete convolution can be defined as:

$$C(i,j) = (I * K)(i,j) = \sum_p \sum_q I(m,n)K(i-p, j-q) \tag{2.14}$$

Discrete convolution can be considered as matrix multiplication. Specifically, the matrix is very sparse because the kernel is usually much smaller than the input image. This property leverages most important ideas in the motivation of CNNs: sparse interactions, parameter sharing and equi-variant representations, which are different from ordinary neural networks. In a deep convolutional network, units in the deeper layers may indirectly interact with a larger portion of the input. This allows the network to efficiently describe complicated interactions between many units. The form of parameter sharing of convolution causes the layer to have a property called equivariance to translation. Invariance to local translation can be a very useful property if we care more about whether some feature is present than exactly where it is.

This property exactly fits our need of feature existence recognition for biological structures. Moreover, by the nature of sparse interactions and parameter sharing, small features in an image can be detected efficiently in terms of the memory requirements and statistics. Details in theoretical explanation can be found in [24], Chapter 9.

Network Architecture

A typical CNN has several layers of transformation: Convolutional Layer, Rectified Linear Unit (ReLU), Pooling Layer and Fully-Connected Layer. An example is shown in Figure 3.4 and Table 3.1. Input layer holds the raw RGB image data as a long one-dimensional vector. Convolutional layer computes the output of neurons corresponding to local regions based on convolution kernels. Each kernel produces one output vector. ReLU layer processes the data by an element-wise activation function: $max(0, x)$ thresholding at zero. This layer does not change the dimensions of data. Pooling layer modifies the output of the net with a summary statistics of the nearby outputs. Simply speaking, it performs a downsampling operation along the vector spatial dimension. Pooling over spatial regions provides invariance to local translation, which is a useful property of preserving features. Fully connected layer's neurons have full connections with all activations in the previous layer, same as those in ordinary neural networks.

The ideal network architecture for a task must be tuned via experimentation guided by monitoring the validation set error. Practically, greater depth does seem to result in better generalization for a wide variety of tasks. Statistically speaking, using deep architectures does express a useful prior over the space of functions the model learns. Three CNN architectures for ImageNet Large-Scale Visual Recognition Challenge (ILSVRC): AlexNet[4], VGG team's ConvNet[25] and GoogleNet[6], have achieved success in this contest. My CNN models will be developed based on these networks. AlexNet (Figure 2.7) introduced the Local Response Normalization (LRN), a desirable property that does not need input normalization to prevent it from saturating. In Figure 2.7, each of the 5 relative large blocks (2nd, 3rd, 4th, 5th and 6th layer) is a combination of convolution, ReLU and LRN operations in order. It is reported that LRN reduces their top-1 and top-5 error rates by 1.4% and 1.2%. AlexNet achieves top-1 and top-5 test set error rates of 37.5% and 17.0% on ILSVRC-2010. VGG team in-

creased the depth of the network by adding more convolutional layers, in their ConvNet (Figure 2.8) models. In ILSVRC-2014, both VGG team and GoogleNet team combined the outputs of several models by averaging their soft-max class posteriors to get ensemble of several networks. In the networks fusion part, GoogleNet (7 nets) was the classification task winner, with 6.7% error, while VGG team (2 nets) secured the 2nd place with 6.8% test error. In terms of the single-net performance, VGG team's architecture achieved the best result (7.0% test error), outperforming the GoogleNet by 0.9%.

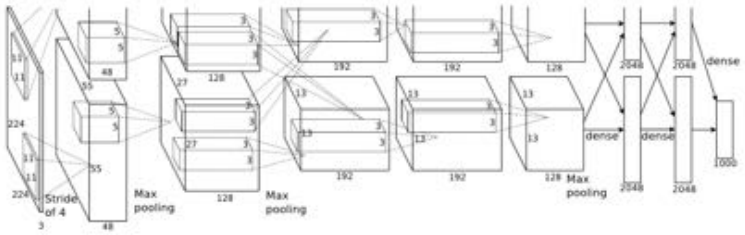


Figure 2.7: An illustration of the architecture of AlexNet[4].

Optimization

CNNs are trained in a purely supervised fashion, using full forward and back-propagation algorithm through the entire network on each training iteration. We cannot use closed formulations to express parameter updates because the nonlinearity of a CNN causes the loss functions to become non-convex. Therefore, CNNs are usually trained by using iterative, gradient-based optimizers. The iterative gradient-based optimization algorithms is widely used in training deep learning networks. For example, caffe, the most popular CNN implementation, uses stochastic gradient descent for optimization. Figure 2.9 shows how to update parameters by gradient descent. Initially, the learning rate ϵ and parameters are required. During the training, the decay of the learning rate may be changed linearly by monitoring the number of iterations. While stochastic gradient descent remains a very popular optimization strategy, learning with it can sometimes be slow. The method of momentum is usually designed in large-scale tasks to accelerate learning. Details in optimization can be found in [24], Chapter 8.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 2.8: VGG team's ConvNet configurations[25].

2.2.4 Gradient Descent

In this section, two optimization methods: gradient descent and expectation maximization will be introduced. Gradient descent is the most widely used first-order optimization algorithm. It works in spaces of any number of dimensions. Expectation Maximization (EM) is a general technique for finding maximum likelihood estimators in latent variable models. Shape priors can be incorporated into EM framework [26] to solve complicated segmentation problems due to color non-standardization.

Gradient descent is a first-order optimization algorithm. For any objective function $f : R^n \to R$, assuming f is convex and differentiable, we want solve the problem: $\min_{X \in R^n} f(X)$ by finding X^* such that $f(X^*) =$

Require: Learning rate ϵ_k.
Require: Initial parameter $\boldsymbol{\theta}$
while stopping criterion not met **do**
Sample a minibatch of m examples from the training set $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ with corresponding targets $\boldsymbol{y}^{(i)}$.
Compute gradient estimate: $\hat{\boldsymbol{g}} \leftarrow +\frac{1}{m}\nabla_{\boldsymbol{\theta}} \sum_i L(f(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}), \boldsymbol{y}^{(i)})$
Apply update: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \epsilon \hat{\boldsymbol{g}}$
end while

Figure 2.9: Stochastic gradient descent (SGD) update at iteration k.

$minf(X)$. Using Gradient Descent, we choose initial $X^{(0)} \in R^n$, repeat:

$$X^{(t)} = X^{(t-1)} - \alpha_t \cdot \nabla f(X^{(t-1)}),\ t = 1, 2, 3, ... \tag{2.15}$$

This iteration can continue until convergence at local or global minimum. Figure 2.10 visualizes the steps of several points from an arbitrary point to the global minimum on a 2D surface. The light blue surface is the objective function f. The red dots are the position $x^{(t)}$ in each iteration. In this example, $f(x, y) = x^2 + 0.01 \cdot y^2$, in which $\alpha_t = 0.1$ and $X^{(0)} = (-6, 13)$.

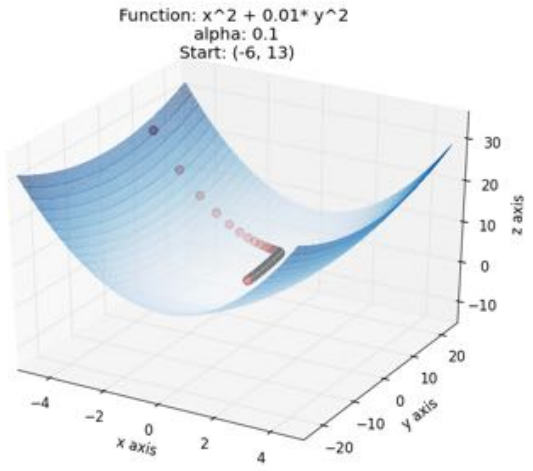


Figure 2.10: The visualization of gradient descent steps on a 2D surface.

Chapter 3

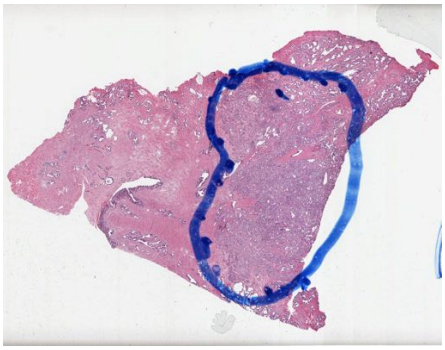
Intermediate Prostate Cancer Classification

We present a fully automatic approach to grade intermediate prostate malignancy with hematoxylin & eosin-stained whole slide images. Deep learning architectures such as convolutional neural networks (CNNs) have been utilized in the domain of histopathology for automated carcinoma detection and classification. However, few work show its power in discriminating intermediate Gleason patterns, partially due to sporadic distribution of prostate glands on stained biopsy samples. We propose optimized staining pigment decomposition on localized images, followed by CNN to classify Gleason patterns 3+4 and 4+3 without handcrafted features or gland segmentation. Crucial glands morphology and structural relationship of nuclei are extracted twice in different color space by the multi-scale strategy to mimic pathologists' visual examination. The presented classification scheme evaluated on 169 whole slide images yielded a 70.41% accuracy and corresponding area under the receiver operating characteristic curve of 0.7247.

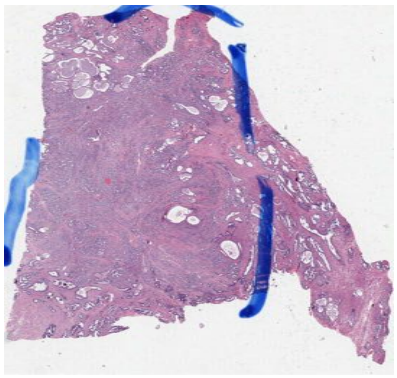
3.1 Introduction

Histopathology is the study of microscopic examination of tissues and cells to uncover the mechanism of disease. The study of histology images was regarded as the reference standard to identify disease for diagnosis and treatment, especially cancer grading [27]. Recently, the digitization of high-resolution whole slide images (WSI) makes it possible to implement computer

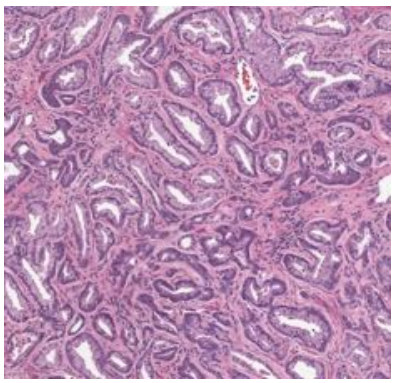
aided diagnosis system to analyze large-scale image data, thus alleviating intra- and inter-observation variations among pathological experts [28]. The computational algorithms and analytical approaches have been developed for cancer detection and grading [29].



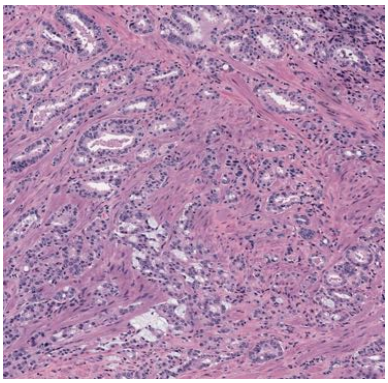
(a) TCGA-EJ-5497 Grade:3+4.



(b) TCGA-EJ-5527 Grade:4+3.



(c) patch of marked area in (a).



(d) patch of marked area in (b).

Figure 3.1: Intermediate prostate cancer image examples.

Prostate cancer shows a favorable long-term prognosis with more than five years of survivor after diagnosis [30]. Gleason score is the grading sys-

tem for prostate cancer based on analyzing patterns of glandular and nuclear morphology [31]. Generally, primary grade is assigned to the dominant pattern of the tumor (with greater than 50% in area seen) and secondary grade is assigned to the subordinate pattern. Each of them has 5 degrees (1-5) according to the extent of carcinoma infiltration, as appearance of recognizable glands. Prostate cancer outcomes, in terms of recurrence risk and specific mortality, differ particularly at the different Gleason 7 patterns (3+4, 4+3) after primary therapy [32]. The 10-year prostate cancer specific survival rate for 3+4 (92.1%) was significantly higher than that for 4+3 (76.5%). Figure 3.1 shows two typical WSI of Gleason 7 patterns (3+4, 4+3) and the patches extracted from them (Figure 3.1(c) and Figure 3.1(d)). The marked area in Figure 3.1(a) and Figure 3.1(b) are target areas evaluated during Gleason grading. We can easily tell that the structure of glands in Grade 3+4 is more complete than that of Grade 4+3. If prostate cancer patients with Grade 4+3 can be distinguished successfully from Grade 3+4, they would receive appropriate treatment and survive longer. So, we are motivated by this clinically meaningful issue to innovate appropriate computer aided diagnosis approach for intermediate malignant prostate cancer grading.

For automatic grading of prostate cancer via histopathological image analysis, many state-of-the-art works have been conducted in gland feature selection [33]. In contrast to human feature engineering, CNNs, as fully data driven approaches, have shown satisfying performance in various computer vision tasks [4]. CNNs are constructed by multiple layers of non-linear transformation and max-pooling, followed by fully-connected layers and a soft-max classifier. ImageNet, a moderate size CNN, contains 60 million parameters and 650,000 neurons. Therefore, with sufficient training data sets and GPU computation resource, CNNs are suitable for histopathology image classification tasks.

In the case of binary classification for prostate Gleason patterns 3+4 and 4+3, inhomogeneous tissue (malignant gland, benign gland, stroma, nuclei, etc.) distributions make pattern recognition less efficient. We intend to delineate irrelevant image information such as benign glands and stroma, so that the training data sets consist only malignant glands, which are the object of Gleason score evaluation. For interest object segmentation, K-Means clustering is a popular algorithm for the segmentation of color images in L*a*b* color space [34]. In hematoxylin & eosin-stained tissue images, the contribu-

tions from each of the applied stains are convolved in RGB color space. To quantify the contribution of hematoxylin, binding to basophilic substances (such as DNA/RNA - which are the main compositions of cell nuclei) [35], a general color de-convolution method [10] based on orthonormal transformation is widely adopted in histopathology analysis. Nevertheless, the separation of two colors comes with an irreversible matrix or an ill-conditioned de-convolution, which makes densities of stains incorrect. In this study, we develop an optimized color decomposition method, for hematoxylin density extraction. The sum of RGB values in the third channel (background) and the changes in H&E channels is minimized using Quasi-Newton method [36]. We applied our method on H&E stained patches extracted from localized areas in WSI. The localized areas, assumed as prostate gland rich areas, are obtained by K-Means clustering. The gray-scale patches doubly extracted from Grade 3+3 and Grade 4+4 WSIs are set as training data set. Classified patches are accumulated to get final accuracy on WSI level. Grade 3+4 and Grade 4+3 will be graded to WSIs with corresponding dominant pattern of the tumor (Grade 3+3 and Grade 4+4).

3.2 Related Work

Substantial histopathology studies on computer-aided grading for prostate cancer investigate novel methods in every aspect of image recognition. Most of them take classical computer vision approach: compute feature information from images, train a classification model, and predict grade using trained model. A number of works propose novel shape descriptors to discriminate gland morphology [37]. Some studies develop classification and regularization method to improve precision [14]. However, most of them focus on distinguishing Gleason patterns 3 and 4, which are less complicated than differentiating Gleason patterns 3+4 and 4+3. Two papers claimed their approaches on automatic grading intermediate prostate cancer [38] [39], but their systems are evaluated with images annotated by expert pathologists or with sub-images from regions of interest, which involve manual data selection. In recent years, convolutional neural networks have been used to detect mitosis in breast cancer histology images [40] [41]. In other histopathology image analysis research, deep learning architectures are applied in basal-cell carcinoma detection [42], brain tumor [43] and invasive ductal carcinoma

[44]. As far as know, this study is the first attempt to use CNN in automatic prostate cancer classification.

3.3 Data

The original whole slide images are downloaded from The Cancer Genome Atlas (TCGA). We use 31 Grade 3+3 WSIs and 34 Grade 4+4 WSIs as our training set. For validation, 3 Grade 3+3 WSIs and 3 Grade 4+4 WSIs are used. In the testing stage, 99 Grade 3+4 WSIs and 70 Grade 4+3 WSIs are utilized for prediction. We use the openslide library to obtain 2-D image data in regions of interest from different layers of WSIs. In localized regions, patches (512 pixels by 512 pixels) are extracted from level 1 with 50% horizontal and vertical overlap. The number of patches extracted from a WSI ("baseline") ranges from hundreds to over 10,000. Our optimized color decomposition method outputs images with 1 channel (hematoxylin), instead of original RGB channels.

We select originally informative patches for "baseline" experiment based on RGB mean values. Five standards are set to exclude patches: 1) RGB mean value higher than 220; 2) or lower than 50; 3) R channel higher than 200 while others lower than 100; 4) B channel higher than 200 while others lower than 100) and 5) image size (less than 200k).

3.4 Methods

In this section, the methods will be introduced according to their orders in the classification pipeline.

3.4.1 K-Means Clustering in L*a*b* Color Space

The localized areas (glands) of WSI for further analysis are segmented by K-Means algorithm in CIE L*a*b* color space. The L*a*b* color space is a 3-axis color system with dimension L for lightness and a and b for the color-opponent dimensions. The L*a*b* color space is based on human perception.

The values in the L*a*b* color space describe all of the colors human could recognize. Since it is in the form of color representation, not the quantity that printing instrumentations need, it is regarded as machine unrelated. L (lightness) ranges from 0 to 100. When the value $L = 50$, it stands for 50% black. The ranges of a and b are both from -128 to +127. In the range, $+127a$ corresponds to pure red, and -128a stands for pure green. Similarly, $+127b$ corresponds to pure yellow, and -128b stands for pure blue. Each color is represented by the combination of the three values.

The values of the two channels (a and b) on each pixel are clustered by K-Means quantization, minimizing the sum of distance functions of each point in the cluster to the K center [45]:

$$\underset{s}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2 \tag{3.1}$$

where $(x_1, x_2, ..., x_n)$ observations are partitioned into $k \leq n$, μ_i is the mean of points in S_i.

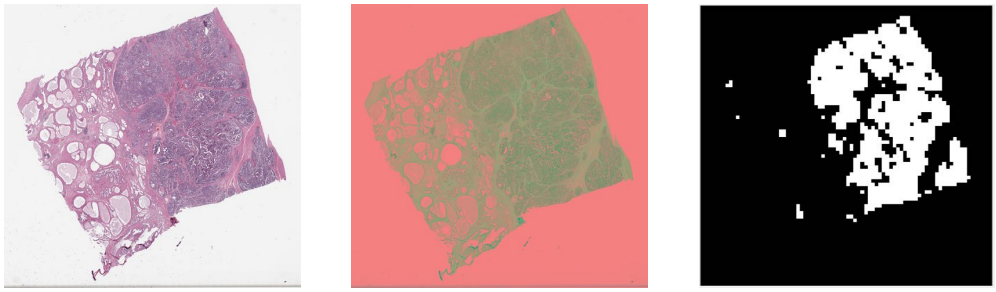
The lowest resolution level of each WSI is converted into L*a*b* color space, as example shown in Figure 3.2 (a) and(b). WSIs without pathologists' mark are clustered into 3 groups, while others are clustered into 4 groups. We select the cluster with second maximum mean value in blue channel. The final binary patch masks on level 1, as shown in Figure 3.2(c), are generated to localize gland areas for further color decomposition and classification.

3.4.2 Optimized Color Decomposition

Original Formulation of Color Decomposition

Given an image $I(x, y, c)$, its optical density (absorbance) is computed as:

$$O(x, y, c) = -\log\left(\frac{I(x, y, c) + 1}{256}\right) \tag{3.2}$$



(a) RGB Space. (b) L*a*b* Space. (c) L1 Patch Mask.

Figure 3.2: K-Means clustering algorithm on WSI.

The three "characteristic" vectors for three channels, $u, v, w \in \mathbb{R}^3$, with $\|u\| = \|v\| = \|w\| = 1$, form a matrix M as

$$M = [u, v, w] \in \mathbb{R}^{3\times 3} \tag{3.3}$$

For each point (x, y), let $R \in \mathbb{R}^3$ as $R = O(x, y, :)$. Then, the decopmosed vector S is $S = M^{-1}R$. Figure 3.3 shows two typical Grade 3+3 and Grade 4+4 patches (Figure 3.3 (a) and (b)) and corresponding results of Hematoxylin channel decomposition (Fig.3.3 (c) and (d)).

Optimization

From Figure 3.3 (c) and (d), we can tell that the density of them differ a lot, due to tissue density variance or staining process variance. The classifier can easily differ these two image patches by their density, but not the gland morphology. This is a strong bias in the system. So we have to manipulate the matrix M to adjust decomposition process for each whole slide image to avoid large systematic color variance. In order to produce a specific matrix M for each whole slide image, we will need to optimize matrix M by the color feature of each whole slide image. Obviously, the third row in decomposed vector S is unrelated to further steps. In original color decomposition method, we just regard them and fix matrix M for each whole slide image. In the optimization part, we propose an idea to minimize the sum of the third row of S and the absolute changes in D. With a multiplier λ , this

optimization process is similar to the second order of regularization.

We can directly optimize the matrix $D := M^{-1}$.
Define

$$E(D) := \int\int \|S(3)\|^2 dxdy + \lambda\|D - \bar{D}\|^2 \tag{3.4}$$

$$= \int\int \|DO(x, y, c)\|^2 dxdy + \lambda\|D - \bar{D}\|^2 \tag{3.5}$$

$$= \int\int (\boldsymbol{d} \cdot O(x, y, c))^2 dxdy + \lambda\|D - \bar{D}\|^2 \tag{3.6}$$

where $\bar{D}$ is the original D matrix from experience and $\boldsymbol{d}$ is the 3rd row of D.

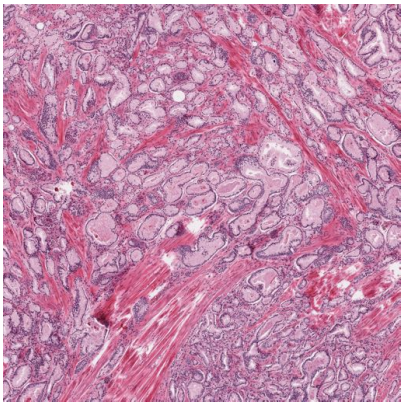
Denote D as

$$D = \begin{bmatrix} D_1 & D_2 & D_3 \\ D_4 & D_5 & D_6 \\ D_7 & D_8 & D_9 \end{bmatrix} \tag{3.7}$$

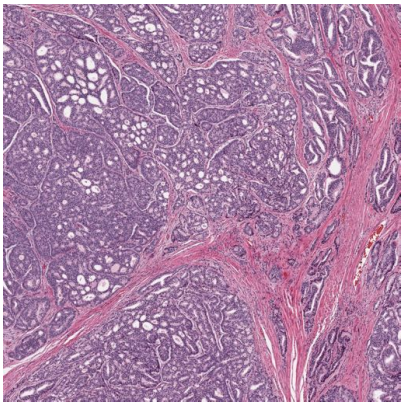
We use the quasi-newton algorithm to minimize $E(D)$ in MATLAB Optimization Toolbox.

3.4.3 Convolutional Neural Network

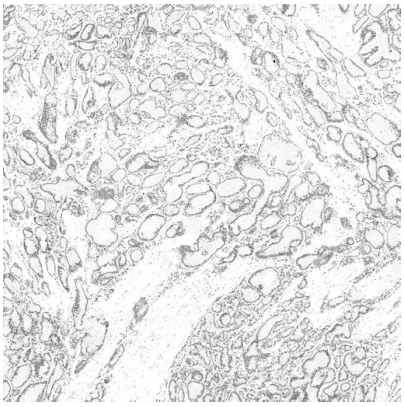
We adopted the CNN architecture from [4], with appropriate modifications for our application. All the layers are constructed as shown in Figure 3.4 [46]. Layers 1, 2 and 5 consist of convolutional (Conv) layers, Rectified Linear Unit (ReLu) activation layers and Max-pooling layers. Layer 6 and 7 are fully connected layers, as in traditional neural networks. The outputs of the Layer 8 are two neurons, representing Grade 3+3 and Grade 4+4. They are activated by a logistic regression model called Softmax. The specification of our CNN architecture is listed in Table 3.1. In the training process, we augment data by setting mirror (reflection of original images). The batch size for training data set is 64. The drop out ratios in Layer 6 and 7 are both 0.5. Each CNN is trained on GPU for 100,000 iterations (approximately 50 hours on one Nvidia K80 GPU).



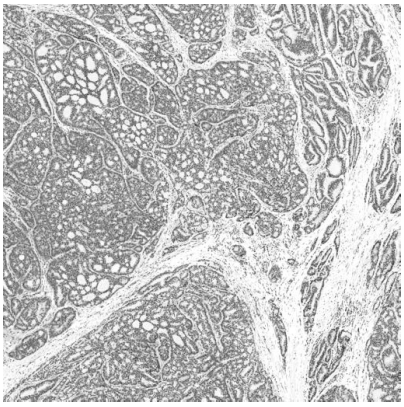
(a) Grade:3+3.



(b) Grade:4+4.



(c) hematoxylin density in (a).



(d) hematoxylin density in (b).

Figure 3.3: The original color decomposition results.

3.5 Results and Discussion

The results of our prostate cancer classification experiments are shown in Table 3.2. We designed six experiments to test the performance of gland localization (3.4.1) and optimization color decomposition (3.4.2). Color decomposition is described in 3.4.2. The convolutional neural networks for each experiment are the same except for the input image channel number. CNNs are trained by Grade 3+3 and Grade 4+4 patches. After prediction, classified patches are accumulated to get final accuracy on WSI level. Grade 3+4 will be graded to those WSIs with more than 50% of the region comprising Grade 3+3. Similarly, Grade 4+3 will be graded to those WSIs with more

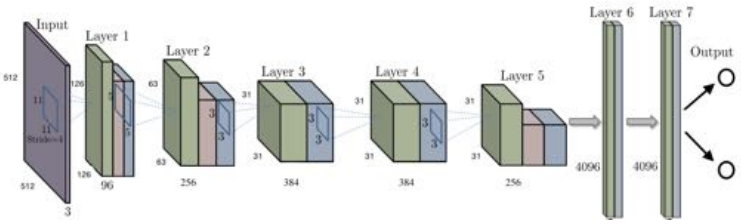


Figure 3.4: The convolutional neural network architecture in our pipeline.

than 50% of the region comprising Grade 4+4. From the table, we can easily observe that only two models: "color decomposition on localized WSIs" and "optimized color decomposition" are able to classify intermediate prostate cancer, with accuracy of 67.46% and 70.41% respectively. Optimized color decomposition has positive effect on prediction performance.

The receiver operating characteristic (ROC) curves for two classification results are plotted in Figure 3.5. By majority, the ROC curve of optimized color decomposition method is on the left-top of that of color decomposition method. Therefore, our proposed optimized color decomposition has general advantage on intermediate prostate cancer automatic grading. The area under ROC curve of optimized color decomposition method is 0.7247, which can be considered as accurate comparable to the first four models listed in Table 3.2. The first four experiment results show no discriminative ability in classifying Grade 3+4 and Grade 4+3, mainly because eosin stained tissue makes training data and testing data no specialty. Structure features of hematoxylin and eosin are totally different, so there is no point to predict prostate cancer based on image patches without appropriate screening.

3.6 Conclusion and Future Works

In this paper, we have introduced deep learning architecture trained by optimized color decomposed image patches in localized areas. The workflow consists of gland region segmentation by K-Means clustering, optimized color decomposition of hematoxylin channel and convolutional neural network classification. The first two steps extract target image information in L*a*b* color space and Euclidean color space respectively. Experimentations have demonstrated that the optimized color decomposition of localized WSIs can

Table 3.1: The architecture parameters of the CNN.

Layers	Filter	Stride	Pad	Output Size
Input	—	—	—	$512 \times 512 \times 3$
Conv	11×11	4	—	$126 \times 126 \times 96$
ReLu+LRN	—	—	—	$126 \times 126 \times 96$
Max-pool	3×3	2	—	$63 \times 63 \times 96$
Conv	5×5	1	2	$63 \times 63 \times 256$
ReLu+LRN	—	—	—	$63 \times 63 \times 256$
Max-pool	3×3	2	—	$31 \times 31 \times 256$
Conv	3×3	1	1	$31 \times 31 \times 384$
ReLu+LRN	—	—	—	$31 \times 31 \times 384$
Conv	3×3	1	1	$31 \times 31 \times 384$
ReLu+LRN	—	—	—	$31 \times 31 \times 384$
Conv	3×3	1	1	$31 \times 31 \times 256$
ReLu+LRN	—	—	—	$31 \times 31 \times 256$
Max-pool	3×3	2	—	$15 \times 15 \times 256$
FC	—	—	—	4096
ReLu+Drop	—	—	—	4096
FC	—	—	—	4096
ReLu+Drop	—	—	—	4096
FC	—	—	—	2
Softmax	—	—	—	2

achieve feasible classification accuracy. Our work provides a new approach for intermediate prostate cancer grading without handcrafted ground truth (as in those approaches mentioned in the related works part), which makes large scale (eg. hundreds of WSIs) processing possible.

Future works can be considered in many promising directions to improve classification precision. For example, more iterations for each model may make loss convergence more satisfying. Moreover, by K-Means clustering, each pixel's distance to cluster center is given, which can be utilized in patch selection. In optimized color decomposition part, ratio λ should be adapted for more accurate prediction. For more practical evaluation, we could invite some medical students or volunteers to manually grade the whole slide images in the test datasets, thus providing baseline from human observers. In this situation, we could add the ROC curves of human observers into Figure 3.5 for more realistic comparisons.

Table 3.2: The prediction results of the experiments.

Experiments	Accuracy
Baseline (Originally Informative Patches)	46.15%
Color Decomposition	53.25%
Optimized Color Decomposition	49.70%
Baseline + K-Means	46.75%
Color Decomposition + K-Means	**67.46%**
Optimized Color Decomposition + K-Means	**70.41%**

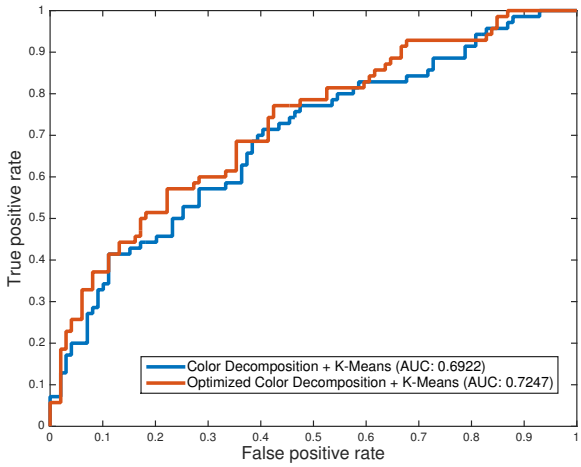


Figure 3.5: The ROC curves for the results of proposed methods.

Moreover, in the respect of practice in the clinic, possible works can be conducted in more realistic cancer prediction. For example, all of ten of prostate cancer scores, in terms of combination of primary score and secondary score in the Gleason system, can be designed to be 10 classes in the classification prediction system. There are a number of whole slide images from each of these 10 classes, while the quantity distribution is not uniform, which means the training data set is imbalanced. Furthermore, the number of whole slide images in some of the classes is very small, so we might come up with a good data augmentation method to make the training feasible in our framework. If we want to learn a deep learning classifier based on these data, we need to tackle these issues at first.

Chapter 4

Nuclei Segmentation for Various Types of Cancer

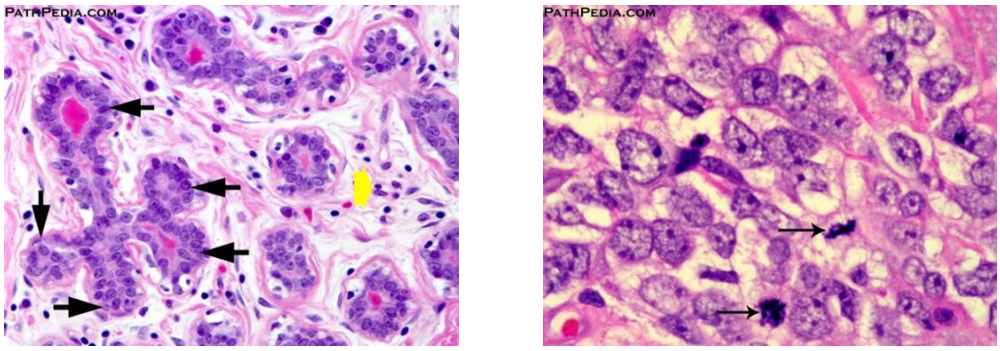
Automatic nuclei segmentation plays an important role in the study of histopathology images. We use a simple yet effective deep learning framework trained with end-to-end pixel-wise supervision for nuclei segmentation in various types of cancer histopathology images. In order to preserve all spatial information, We propose to use no pooling layers nor strided convolutions. Additionally, to separate clustered nuclei more accurately, we enhance the edge representation capability of the deep convolutional networks by hard negative mining. Our approach is fully automatic without pre-, post-processing nor fine-tuned hyper-parameters. We evaluate our method on the 2017 MICCAI Nuclei Segmentation Challenge dataset [47] and outperform the challenge-winning result.

4.1 Introduction

4.1.1 Background and Motivation

Histopathology is the branch of medicine associated with the study of the essence of disease and its causes, processes, development and consequences [48]. In the most commonly used Hematoxylin and Eosin stained histopathology images, Hematoxylin stains nucleic acids (nuclear chromatin and nucleoli) blue, whereas Eosin stains proteins (cytoplasm and connective tissue)

pink [49]. Figure 4.1 shows two Hematoxylin and Eosin (H&E) stained breast tissue images. We can see the nuclei are stained blue and we can even distinguish the nucleoli and nuclear matrix within nucleus.



(a) Normal (b) Invasive Ductal Carcinoma

Figure 4.1: Examples of H&E breast tssues [50]. The morphology, appearance, distribution of nuclei (indicated by arrows) distinguish normal versus cancerous tissues.

Nuclear atypia, or nuclear pleomorphism is usually the diagnostic standard of cancer grading. Pleomorphism is commonly the variation of size in malignant nuclei. In Figure 4.1, we can easily tell the uniformity difference of nuclei in normal and invasive ductal carcinoma breast tissues. The shape of the nucleus, frequently abnormal in carcinoma cells (Figure 4.1(b)), is dictated by the characteristics of the nuclear margin. Specifically, salient visual morphological features of cancer nuclei in H&E images are nuclear margin, nucleoli, nuclear chromatin and nuclear matrix. For nuclear envelope, or nuclear margin, the nuclear boundary is regularly rounded, oval and smooth in normal cells [51], while in malignancy, the nuclei become angulated with irregular thickening of nuclear margin and often-nuclear moulding, as we can see those relatively uniform oval nuclei in Figure 4.1(a) and irregular shape distribution in the nuclei in Figure 4.1(b). Furthermore, the morphological alterations of cancer nuclei are also likely to be associated with larger nucleolus of the proliferating cells, changes of chromatin pattern (the peripheral heterochromatin formation in the nuclear periphery of the malignant nuclei) and vesicular nucleus with peripheral margination of nuclear matrix [52]. In Figure 4.1(a) and (b), nucleolus size variation, chromatin patterns and nu-

clear matrix density difference could all be discerned clearly.

The morphological features of neoplastic cell nuclei are important in the diagnosis and grading of various forms of neoplasms [52]. In the field of image analysis, a lot of researchers are utilizing effective machine learning and deep learning algorithms to diagnose cancer automatically. In order to do objective analysis of nuclear morphology features, we need to obtain nuclei boundaries at first. Therefore, fully automatic nuclei segmentation methods of H&E images are essential prerequisite to automatic computer aided diagnosis of different types of cancer. Recently, convolutional neural network (CNN) has been the most powerful method in semantic segmentation [53]. The efficiency of deep learning is the capability to learn appropriate feature representations for the specific problem in an end-to-end fashion. It does not need domain knowledge in hand-crafted features extraction, neither too much fine-tuning in model training. Taking advantages of the availability of high resolution digital pathology images, efficient deep learning algorithms and GPU computation power, automatic nuclei segmentation algorithms could be developed to assist pathologists' diagnosis or further analysis.

4.1.2 Related Work

In the past few years, many researchers in pathological image analysis field have explored the capabilities of convolutional neural networks for nuclei segmentation. In [54], the authors proposed a multi-scale convolutional network scheme to segment cervical cytoplasm and nuclei. After deep learning coarse segmentation, a graph partitioning method was developed with learned feature to localize the appearance of distinctive boundary. Xing [55] proposed a learning-based framework for robust and automatic nucleus segmentation of brain tumor, pancreatic neuroendocrine tumor, and breast cancer. Those two paper both have two stages, with deep learning performing coarse segmentation, followed by traditional feature based segmentation. It starts with a deep convolutional network with an iterative region merging approach for contour initialization. Then, a more accurate segmentation algorithm is exploited based on repulsive deformable model constrained by sparsity-based shape prior. The approach in [56] augmented challenging patches sampling. The process of hard negative mining improved the nuclei boundary delineation for estrogen receptor positive breast cancer. The framework of [57] could be divided into three stages: 1. background removal by the sparse

reconstruction method; 2. deep network segmentation; 3. post-process with morphological operations and prior knowledge. The deep contour-aware networks [58] improved fully convolutional network to segment nuclei and glands from histology images. For better separation the touching nuclei, a multi-task learning framework was designed to predict objects and contours individually. The CAE (Convolutional Autoencoder) [59] is a semi-supervised learning approach, which detects and encodes nuclei based on sparse feature maps that encode both the location and appearance of nuclei.

4.1.3 Overview

Those contributions to nuclei segmentation have obtained reasonable results, however, they require pre- or post-processing and/or data specific processing. We use a simpler yet effective end-to-end deep learning scheme with edge representation and pixel-wise supervision. **Our main contribution** is to use a deeper convolutional network without pooling layer. Additionally, to separate clustered nuclei more accurately, we enhance the edge representation capability of the deep convolutional networks by hard negative mining. We evaluate our method on the 2017 MICCAI Nuclei Segmentation Challenge, our method achieves the average dice coefficient of 0.7847, outperforming the challenge winning team. By employing transfer learning and ensemble techniques, our results reach the score of 0.7871 with an individual model and 0.7893 with an ensemble model.

We compare our method with those six methods introduced in the related work section. Using our individual model result which is without transfer learning, the dice-1 is 0.8560, and the dice-2 is 0.7133, thus with the average of 0.7847. In [54], the dice-1 reached 0.912 on cervical nuclei. Although the results are really good, this approach is not fully based on deep learning. Before CNN, it uses texture, shape and contextual information of the objects to segment nuclei by graph partitioning method. Our end-to-end deep learning method is an integrated approach, which is simpler and more robust to future data. In [55], the authors use a CNN to generate a probability map, and then use an iterative region merging approach and to generate shape initialization and finally, use a sparse shape and deformable combined model to separate individual nuclei. This method obtained the average dice-1 of 0.77, 0.88 and 0.78 on brain tumor, pancreatic neuroendocrine tumor and breast cancer respectively. This method applies two very complicated seg-

mentation methods after deep learning, so it needs many domain knowledge, which cannot the simplicity of our method. In [56], this method emphasizes the edge information in deep learning approach, reaching the dice-1 of 0.83 across 12,000 breast cancer nuclei. I borrow the negative edge mining to apply in my approach. However, instead of using single label for each patch, I use pixel-wise patch training scheme like bag algorithm to improve the model performance. Therefore, in similar amount and type of cancer, our method is marginally higher than this method by 0.026 in respect of dice-1. In [57], this paper applies a sparse reconstruction method, CNN, and morphological operations to improve the segmentation performance. On 58 breast cancer images, the method reaches dice-1 of 0.8393. This method employs pre-processing and post-processing besides deep learning, thus needing prior knowledge. In [58], the authors propose a contour-aware network for histopathological object segmentation. They use two network to output nuclei and also its boundaries. Their obtained the dice-1 of 0.876, dice-2 of 0.748 and the average of 0.812, based on the dataset of 2015 MICCAI Nuclei Segmentation Challenge. Although their dices are higher than our method, their data are different to what we used. [59] is a semi-supervised approach of nuclei detection and segmentation. It uses a sparse convolutional auto-encoder (CAE) to extract feature and a CNN to classify patches. The dice-1 reaches 0.8362 on the dataset of 2015 MICCAI Nuclei Segmentation Challenge. Different from our fully supervised integrated approach, it need to pre-train the CAE as a feature extractor.

4.2 Methods

For the model architecture, we design two types of convolutional network architectures, each with three kinds of depths. Different from other segmentation deep learning approaches, we do not apply any pooling layers in our architecture. During training, we design the label to be a patch with the same size of the training patches, which intelligently improves the nuclei boundary discrimination capability. In regard to batch sampling, we borrow the idea from [56], augmenting the challenging negative patches with the appropriate ratio.

4.2.1 Deep CNN Models

We design two main CNN network models: the straight one (Figure 4.2) and the concatenated one (Figure 4.3). For each mode, we exploit three different depths to try to capture more nuclei features. For the straight models, the architecture of the 3 straight models are listed in Table 4.1. (16-layer, 23-layer and 30-layer). Their convolutional layer filter size are same in each block. For example, the16-layer model has 5 convolutional blocks and 1 fully connection block. We apply leaky rectify nonlinear unit after each layer as activation function. For the concatenated models, we combine different blocks of feature maps to obtain more information. The 16-layer concatenation model is shown in Figure 4.3. Similarly, the 23-layer concatenation model combines Layer 5 & Layer 10, and Layer 13 & Input. The 30-layer concatenation model combines Layer 7 & Layer 14, and Layer 20 & Input.

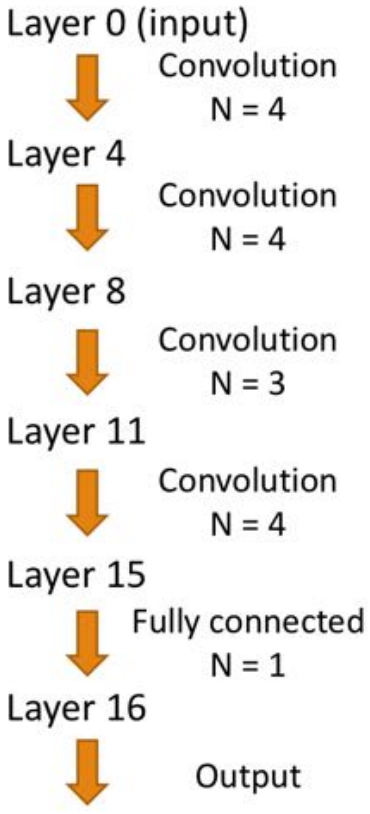


Figure 4.2: The computation flow of the 16-layer straight model.

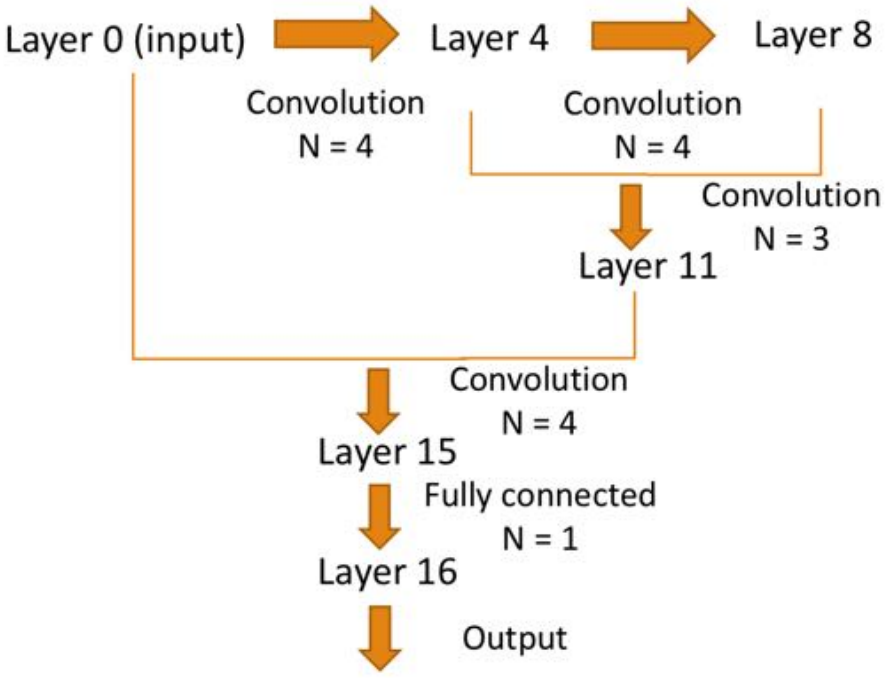


Figure 4.3: The computation flow of the 16-layer concatenated model.

4.2.2 Novel Pixel-wise Label

A large amount of CNN based methods follow a sliding window scheme. The sliding window is trained and predicted at the center where the pixel is labeled and to be regressed. In our framework, the size of the training patches is 40×40. We also designed the labels to be 40×40 patches, which leverage the ability of deep representation to improve the nuclei boundary discrimination. Furthermore, this technique is equivalent to training 40×40 models simultaneously. In each training iteration, we randomly select patches, like the random one within the bounding box in the left image shown in Figure 4.4. The enlarged patch is illustrated on the top right image, with its label patch illustrated on the bottom right image.

4.2.3 Negative Edge Mining

Generally, when solving machine learning or deep learning problems, we should keep training datasets distribution the same as that of testing datasets.

Table 4.1: The architecture of three CNN straight models.

Model	**Layers**	**Filter Size**	**Filter Num**
16-layer	Input	$5 \times 5(conv)$	80
	Layer 4	$3 \times 3(conv)$	100
	Layer 9	$3 \times 3(conv)$	64
	Layer 13	$1 \times 1(conv)$	64
	Layer 14	$1 \times 1(conv)$	32
	Layer 15	$1 \times 1(fully)$	1
	Output	sigmoid	—
23-layer	Input	$5 \times 5(conv)$	80
	Layer 5	$3 \times 3(conv)$	100
	Layer 10	$3 \times 3(conv)$	64
	Layer 15	$3 \times 3(conv)$	32
	Layer 20	$1 \times 1(conv)$	32
	Layer 22	$1 \times 1(fully)$	1
	Output	sigmoid	—
30-layer	Input	$5 \times 5(conv)$	80
	Layer 7	$3 \times 3(conv)$	100
	Layer 14	$3 \times 3(conv)$	64
	Layer 22	$3 \times 3(conv)$	32
	Layer 27	$1 \times 1(conv)$	32
	Layer 29	$1 \times 1(fully)$	1
	Output	sigmoid	—

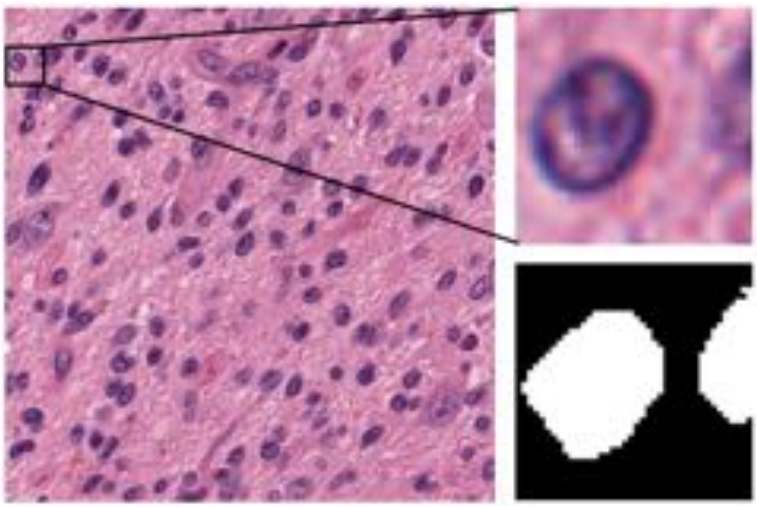


Figure 4.4: Random patch selection and pixel-wise label.

Using stochastic sampling can modeling the training data distribution, however, in the case of nuclei segmentation, random patch sampling usually does not guarantee good prediction performance. Because the edge feature may be underrepresented in the training sets by deep learning models. Therefore, we intend to use the successful idea in [56], in which, the author supplemented the training set with the patches centering on edges of the nuclei.

We also use dilation operation to get the edge masks with three pixels dilated from the binary mask. Specifically, binary image dilation is a process that changing a background pixel to foreground when the background pixel has a 4-neighbor foreground pixel. So, a three-pixel-width nuclei margin binary mask is obtained via processing the binary mask with dilation operation by three iterations, and subtracting the binary mask itself from the dilated binary mask. In Figure 4.5, the left image is the original RGB image. The middle image is the nuclei binary mask of the left one. The right image is the three-pixel-width nuclei margin binary mask of the left one.

The three-pixel-width nuclei margin binary mask (right one in Figure 4.5) demonstrates the negative areas, which also are most near to the edges of the nuclei. During training, if the center of a random selected patch is in the positive area of nuclei margin binary mask, we supplement 12 such patches

into each batch of 80 random patches.

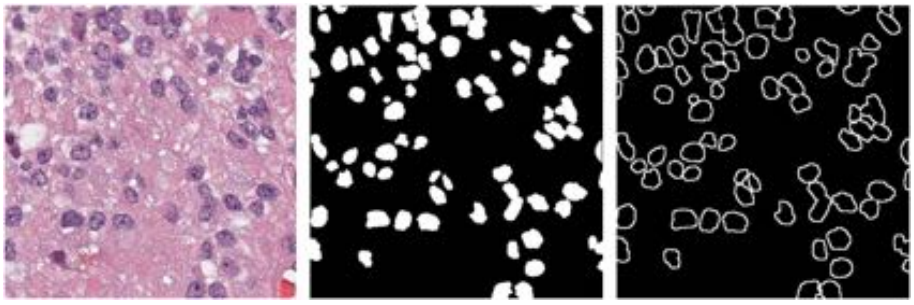


Figure 4.5: The original image, mask and edge mask.

4.3 Evaluation and Datasets

4.3.1 Evaluation Dice Coefficients

In this paper, the proposed method is applied on the datasets from 2017 MICCAI Nuclei Segmentation Challenge[47]. In this challenge, the scoring is composed by one traditional dice coefficient (dice-1) and one variant dice coefficient (dice-2). The dice-1 is the traditional dice coefficient to measure the overall overlapping between the ground truth and the participant segmentation. The dice-2 captures disparity in the way the segmentation connected components are separated, while the overall overlapping may be very promising. The two dice coefficients will be computed for each image tile in the testing dataset. The two dice coefficients for the entire testing dataset will be the average of the scores for each image tile. The average of two dice coefficients will be ranked in the challenge leaderboard.

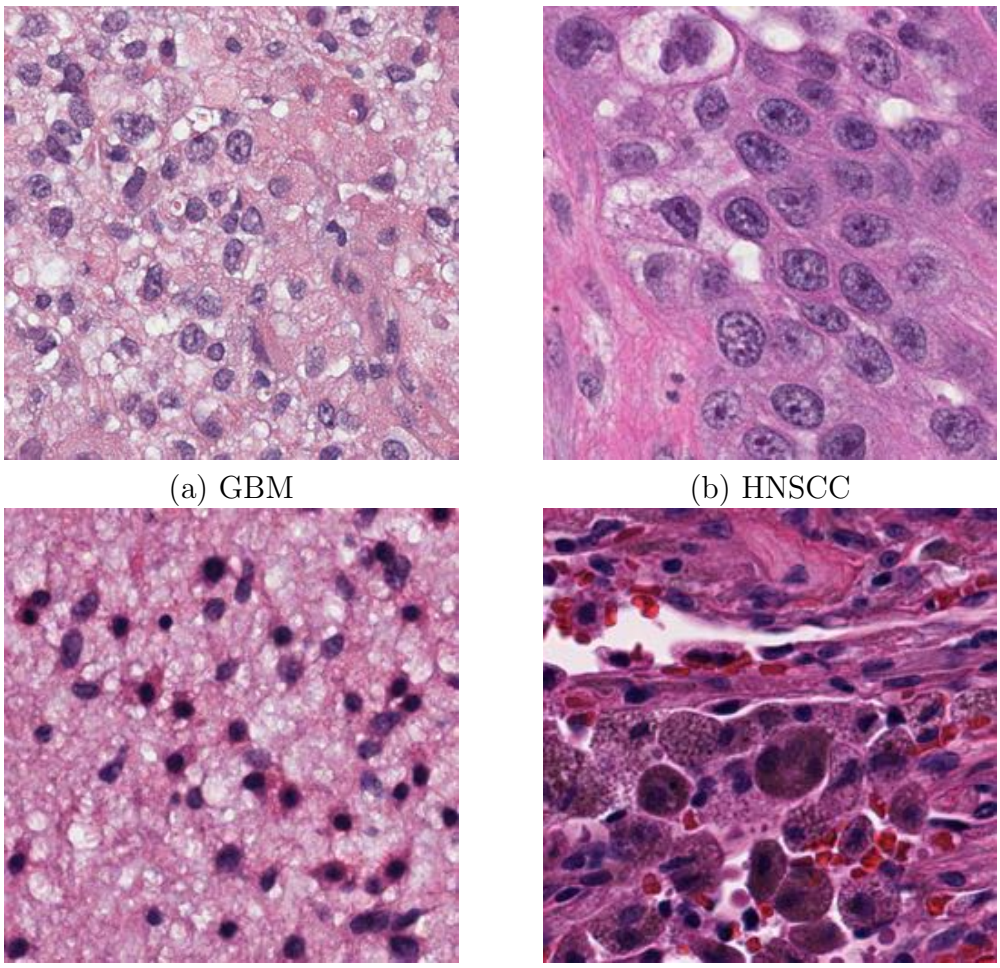
To set up the dices computation algorithms, the inputs are two sets of images. In the ground truth image set, each image is $G_i : \Omega \to \{0\} \cup \mathbb{Z}^+$. In the prediction image set, each image is $I_i : \Omega \subset \mathbb{R}^2 \to \{0, 1\}$. The pseudo-codes (using matlab functions) for dice-1 and dice-2 computation are illustrated below ($J = bwlabel(I, 4)$).

```
DICE-1 COMPUTATION(G, I)
1  sum_G, sum_I, sum_overlap = 0
2  for i ← 1 to length[G]
3        do nnz_G ← nnz(G(i))
4           nnz_I ← nnz(I(i))
5           nnz_overlap ← nnz(G(i)&I(i))
6           sum_G += nnz_G
7           sum_I += nnz_I
8           sum_overlap += nnz_overlap
9  dice − 1 = 2 * sum_overlap/(sum_I + sum_G)
```

```
DICE-2 COMPUTATION(G, J)
 1  sum_G, sum_I, sum_overlap = 0
 2  for i ← 1 to length[G]
 3        do p ← max(G(i))
 4           q ← max(J(i))
 5           for m ← 1 to p
 6                 do s1 = (G(i) == m)
 7                    for n ← 1 to q
 8                          do s2 = (J(i) == q)
 9                             nnz_G ← nnz(s1)
10                             nnz_I ← nnz(s2)
11                             nnz_overlap ← nnz(s1&s2)
12                             sum_G += nnz_G
13                             sum_I += nnz_I
14                             sum_overlap += nnz_overlap
15  dice − 2 = 2 * sum_overlap/(sum_I + sum_G + eps)
```

4.3.2 Datasets

2017 MICCAI Nuclei Segmentation Challenge delivered training and testing images of four types of cancer: 1. glioblastoma multiforme (GBM); 2. head and neck squamous cell carcinoma (HNSCC); 3. lower grade glioma (LGG) tumors; 4. non small cell lung cancer (NSCLC). The training and testing set both consist of 32 image tiles, 8 tiles from each cancer type. Most of the images are of size 500×500. Only a few of the images are of size 600×600. Four example images (as in Figure 4.6) are selected from the testing set. Each of them is from one type of cancer respectively.



(a) GBM (b) HNSCC

(c) LGG (d) NSCLC

Figure 4.6: Testing dataset example images.

4.4 Experiments, Results and Discussion

4.4.1 Implementation

The framework is implemented with python 2.7. We choose theano and lasagne as our deep learning library. All the training and predicting experiments run on 2 Tesla-K80 GPUs.

Training Stage

In the training stage, we use cross entropy to compute loss function and the nesterov_momentum as our parameter update algorithm. Batch normalization is applied after convolutional operation. The batch size is 92 for each iteration. We train each model by 30,000 epochs. Data augmentation is applied via resizing the original images; flipping, rotating and changing the hue of randomly selected patches. Since we also have data from the same challenge in 2015, we merge two years' data to try to enhance feature database.

Testing Stage

In the testing stage, we predict sliding patches in the testing images with a stride of five. By averaging overlapped patches of nuclear probabilities, we can get nuclear probability maps with pixels in the range of zero to one. After transforming probability maps to 8-bit depth of integer images, we can binarize them by an empirical threshold (from 2015 Challenge data) to get final prediction binary maps. Last but not least, we apply the ensemble method of averaging results of six models to improve prediction accuracy. We simply average corresponding probability maps, which is actually a procedure of noise removal to some extent.

4.4.2 Results and Discussion

For each model we design in the method section, we deploy the trained model to get binary nuclei mask for each testing image.

The qualitative segmentation results are shown in Figure 4.7. The yellow(ground truth) and cyan(our segmentation results) boundaries are obtained by erosion operation on nuclei regions. We can compare the deep learning output with ground truth in Figure 4.7. For the GBM, HNSCC and

LGG examples, our method outputs good results with a little weakness in clustered nuclei separation. For the NSCLC example, some of the nuclei are very difficult to capture their boundary appropriately, because the intensity contrasts of the nuclei with that of the cytoplasm are too small.

More objectively, the quantitative segmentation results are shown in Table 4.2. The first half of the table shows dice-1, dice-2 and average of two dices from the six models trained with the dataset from the challenge in 2017. The second half of the table shows the results from the same models, but trained with two datasets merged from the challenge in 2015 and 2017. From the table, five experiments output excellent results (bold) that beat the first ranked average dice (0.783) in the challenge. In general, deeper models produce better results, both in dice-1 and dice-2. Nevertheless, my concatenated models do not show effective progress. Adding the dataset from 2015 also helps improve both scores. Furthermore, two ensemble results are the two best ones in the table, which means ensemble method really enhance the segmentation performance. Obviously, any increase in dice-2 with no trade-off in dice-1 will definitely improve the result, so the de-clumping technique will be an important research project in future.

Although more convolutional layers produce more feature maps (which improves the discriminative ability), the improvement effect depends on the amount of training data and the complexity of the network. Within a fixed network model, the model starts from under-fitting, to best fitting and ends at overfitting along with the increase of the convolutional layer number. We need to avoid overfitting in model design by doing experiments of different layers of convolution. Furthermore, large number of convolutional layers will also cause more computational time, more GPU memory usage and more local minima in the high dimensional spaces. So, we list the computation time in the training stage for each model in Table 4.2. In order to investigate the positive and negative effects of adding layers to the CNN, we add more experiments to the project, by keeping adding layers until the test error does not improve anymore. The additive experiments results are shown in the third part of Table 4.2. In the third part of Table 4.2, based on the prediction accuracy and training time of straight_30_layer, straight_37_layer and straight_44_layer model, we can conclude that keeping adding layers causes much more training time cost, but gains little accuracy improvement. Therefore, with considerable training time, the straight_30_layer model is the most

appropriate model for this nuclei segmentation problem.

Each dataset usually has one set of manual segmentation from pathologists. The manual segmentation from pathologists are regarded as ground truth. Therefore, currently, we may not be able to compute dice coefficients of pathologists' segmentation because we cannot obtain more pathologists' manual segmentation on the same histopathology images, at least no second manual segmentation from pathologists could be found on histopathology online database. Once we find second manual segmentation from pathologists, we will apply dice coefficients on them to evaluate the pathologists work. Based on dice values from section 4.1.3, 2nd paragraph, a dice-1 of over 0.83 is a publishable good result of a segmentation algorithm. Nevertheless, dice value not only depends on segmentation algorithm, but also on image data and segmentation difficulty. Clinically, we compare two sets of segmentation from automatic algorithms by applying a statistical hypothesis test: t-test [60].

4.5 Conclusion

This paper presents a fully automatic end-to-end deep learning framework for nuclei segmentation for various types of cancer. Our method outperforms existing state-of-the-art nuclei segmentation methods.

Our network uses no pooling nor strided convolutional layers, and enhances the edge representation capability of the deep convolutional networks by hard negative mining and pixel-wise supervision. This simple yet effective nuclei segmentation method is deployed without pre-processing, post-processing nor specific domain knowledge.

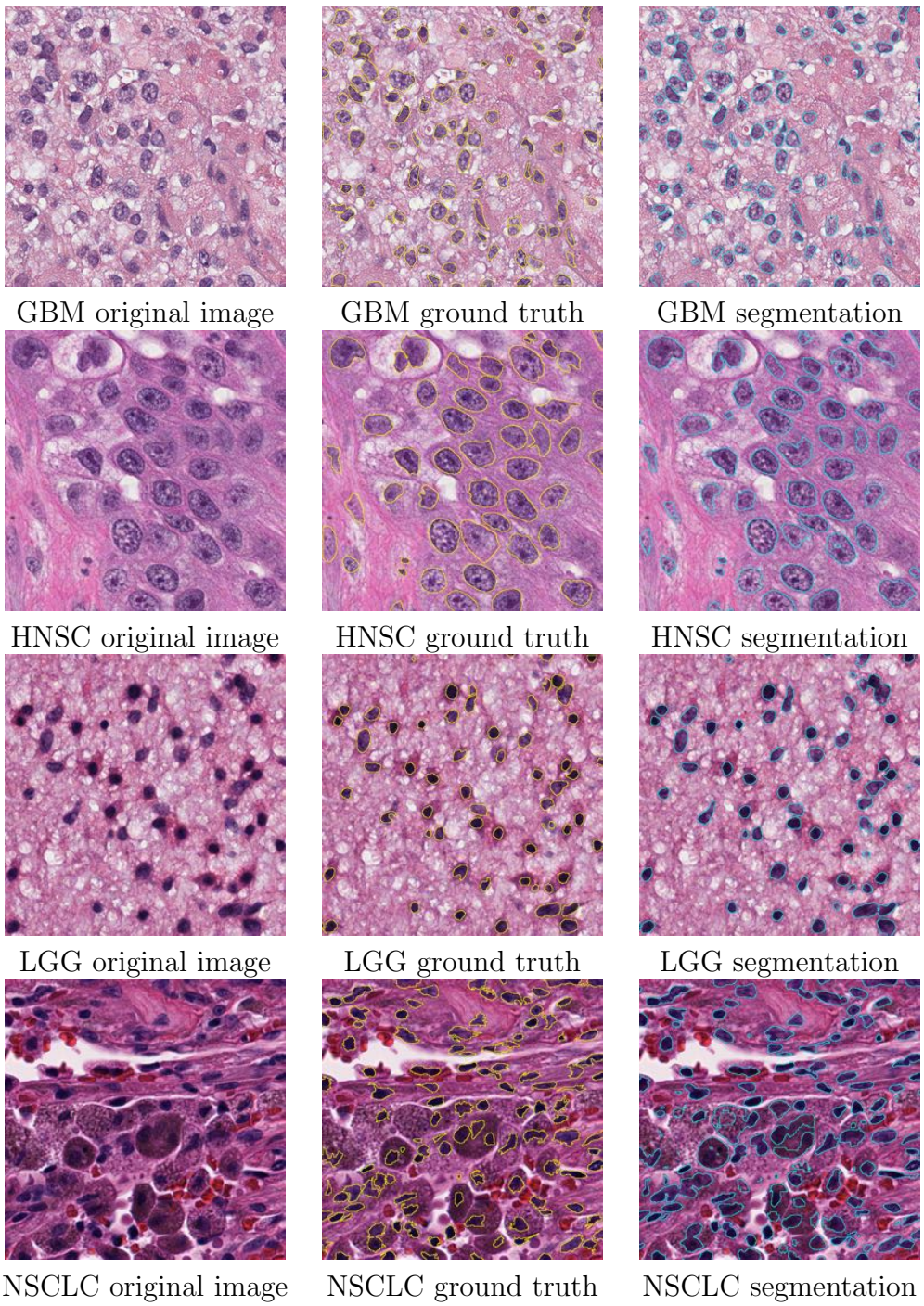


Figure 4.7: The visualization of segmentation results. (left: original images; middle: ground truth; right: our segmentation results)

Table 4.2: The quantitative segmentation results.

Model (2017 data)	**dice-1**	**dice-2**	**Average**	**Time**
straight_16_layer	0.8497	0.6976	0.7737	23 hrs
concatenated_16_layer	0.8459	0.6966	0.7713	28 hrs
straight_23_layer	0.8583	0.6904	0.7744	31 hrs
concatenated_23_layer	0.8593	0.6949	0.7771	34 hrs
straight_30_layer	0.8560	0.7133	**0.7847**	52 hrs
concatenated_30_layer	0.8593	0.6979	0.7786	48 hrs
ensemble	0.8621	0.7132	**0.7877**	
Model (2017+2015 data)	**dice-1**	**dice-2**	**Average**	**Time**
straight_16_layer	0.8544	0.6989	0.7767	25 hrs
concatenated_16_layer	0.8587	0.6920	0.7754	31 hrs
straight_23_layer	0.8615	0.7037	0.7826	33 hrs
concatenated_23_layer	0.8598	0.6859	0.7729	38 hrs
straight_30_layer	0.8567	0.7102	**0.7835**	48 hrs
concatenated_30_layer	0.8614	0.7127	**0.7871**	46 hrs
ensemble	0.8634	0.7152	**0.7893**	
straight_37_layer	0.8592	0.7088	**0.784**	59 hrs
straight_44_layer	0.8636	0.6967	0.7801	81 hrs

Chapter 5

Evaluation of Nuclei Segmentation via Image Synthesis

Digital histopathology images with more than 1 Gigapixel are drawing more and more attention in clinical practice, biomedical research, and computer vision. Among the multiple observable features spanning multiple scales in the pathology images, nuclear morphology is one of the central criteria for diagnosis and grading. As a result it is also the most studied target in image computing. Many research papers have been devoted to the problem of extracting nuclei from digital pathology images, which lies at the foundation of any further correlation study. However, the validation and evaluation of nucleus extraction have yet been formulated rigorously and systematically. Some researcher report a human verified segmentation with thousands of nuclei, whereas a single whole slide image may contain up to million. The main obstacle lies in the difficulty of obtaining such a large number of validated nuclei, which is essentially an impossible task for pathologist. We propose a systematic validation and evaluation approach based on large scale image synthesis. This could facilitate a more quantitatively validated study for current and future histopathology images analysis field.

5.1 Introduction

Digital histopathology image computing is becoming an ever more active field due to its superb spatial resolution and the availability of large data sets. At a spatial scale between genomics and radiology, histopathology images have shown the promising possibility of revealing detailed image based features that can be linked with genetics in a straightforward manner, as well as the epigenetic and environmental factors, which are not present in the DNA sequences.

Among the vast amount of image features in the digital pathology images, the nuclei lie at a central position. Their morphologies are the main criteria for the pathologist making diagnostic and grading decisions. Image computing often starts with nucleus analysis, the accurate extraction of every nuclei being the pre-requisite of such analysis [3].

There exists an extensive literature discussing nucleus segmentation [61, 62, 63, 64, 65, 66, 15, 67, 68, 69, 70]. The quantitative validation and evaluation of a given nucleus segmentation algorithm is one of the most challenging issues in all of the aforementioned publication. Indeed, it is difficult, if not impossible, to ask a skilled pathologist to manually contour thousands of nuclei to be compared with algorithm output. Even so, one whole slide image may contain up to a million nuclei. Therefore, instead of relying on human annotation, we need a systematic approach to computationally generate large enough data set to test the accuracy and robustness of the algorithm, which is the main topic of this paper.

5.2 Methods

5.2.1 Image Synthesis and Quantitative Evaluation Framework

To set the stage: The algorithm starts with a set of images $I_i : \Omega \subset \mathbb{R}^2 \rightarrow C; i = 1, \ldots, N$ where C denotes the RGB color space. We have expert validated segmentation of the nuclei in I_i's, denoted as label images $J_i : \Omega \rightarrow \{0\} \cup \mathbb{Z}^+$. In the label images, 0 denote background and different

integer values indicate different nuclei. Denote the bounding box of the j-th nucleus in I_i as $B_i^j := (x_i^j, y_i^j, u_i^j, v_i^j); j = 1, \ldots, N_i$ where (x, y) (resp. (u, v)) denotes the top-left (resp. bottom-right) coordinates for the bounding box, and N_i is the number of nuclei in I_i. With slight abuse of notation, the image domain of the bounding box is also denoted as B_i^j. Furthermore, the restriction of the image and the label image, i.e., the patch around the nucleus, are denoted as $I_i^j := I_{i|B_i^j}$ and $J_i^j := J_{i|B_i^j}$, respectively. Finally, the non-nuclear region in the patch is removed by defining $I_i^j := I_i^j * J_i^j$. To simplify the notation, we use single index and denote $P_1 = I_i^1$, $P_2 = I_i^2$, and $Q_1 = J_i^1$, $Q_2 = J_i^2$ etc.

Nucleus Synthesis

In this step, we synthesize new nuclei based on those in the training images. To proceed, a random nucleus P is picked from all P_i's. Together with its mask Q, they will be used to generate a new nucleus. In order to generate nuclei with different orientation and thus evaluate the *orientation invariance* of the segmentation algorithm, the patch P and Q are rotated with a random degree $\theta \in [0, 2\pi)$. The size of nucleus is also a key factor being considered in clinical practice. To simulate the *size variation*, the sizes of the training nuclei are computed and the size distribution, p_S, is estimated through a kernel density estimation process [71]. Then, a sample is drawn from p_S and the new nucleus is scaled to the size [72].

After that, a nonlinear deformation is applied to the new nucleus. At the same time, the corresponding label map is also deformed. As a result, we still have the exact segmentation of the nucleus. To this end, one possible approach is to generate a random deformation field and smooth it to certain degree. Such a random deformation may be feasible from a computational point of view. Yet, the random deformation may generate some unrealistic shape. Because of this, we propose a relative conservative approach so that the resulting deformed nucleus is within the space spanned by the training nuclei.

Specifically, for the training nuclei, without loss of generality, we assume their sizes (areas) have been normalized. Then, a Procrustes alignment is performed so that all major axes are aligned north-south. After that,

each nucleus is mapped to the unit disk though the optimal mass transport (OMT) [73, 74]. Formally, a set of points $G := \{\boldsymbol{x}_i \in \mathbb{R}^2 : i = 1, \ldots, m\}$ is sampled uniformly from a unit disk, and a same number of points $H := \{\boldsymbol{y}_i \in \mathbb{R}^2 : i = 1, \ldots, m\}$ are sampled from the nucleus to be registered to the disk. Each point is considered to have a Dirac metric. Then, an optimal matching between the two sets of points is constructed. To that end, we denote the correspondence between X and Y by a matrix $A \in \{0, 1\}^{m \times m}$ where $A_{i,j} = 1(0)$ indicates $\boldsymbol{x}_i$ is corresponding (not corresponding, resp.) with $\boldsymbol{y}_j$. Denoting the pair-wise distance matrix $C \in \mathbb{R}^{m \times m}$ as $C_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{y}_j\|_2$ where $\|\cdot\|_2$ is the L_2 norm, we find the correspondence between the two sets of points by solving such an assignment problem:

$$\begin{aligned} A = & \min_{\tilde{A} \in \mathbb{R}^{m \times m}} \|C \circ \tilde{A}\|_F \\ \text{s.t.} \ & \sum_j \tilde{A}_{i,j} = 1 \quad \forall i \in \{1, \ldots, m\} \\ & \sum_i \tilde{A}_{i,j} = 1 \quad \forall j \in \{1, \ldots, m\} \\ & \tilde{A}_{i,j} \geq 0 \quad \forall i, j \in \{1, \ldots, m\} \end{aligned} \tag{5.1}$$

where $\circ$ is the Hadamard product of the two matrices and $\|\cdot\|_F$ is the matrix Frobenius norm. Moreover, it is noted that the optimization variable $\tilde{A}$ is not restricted to be a binary matrix. Otherwise the optimization becomes an NP-hard combinatorial problem. On the other hand, due to the fact that the constraint matrix of (6.10) is totally unimodular, the resulting optimal A is a binary matrix [75]. This optimization problem can be shown to be convex, and it can be effectively solved by using, for example, the interior point method [76]. The resulting matrix A will give a one-to-one correspondence between G and H. The deformation field D_i is therefore constructed as the displacement vector field among the corresponding points.

In a high dimensional deformation space, the D_i's most likely reside on a manifold rather than in a linear space [77, 78]. In order to generate arbitrary deformations in a similar fashion to that has been observed in the training data, one can interpolate the deformation fields on the manifold. However, due the high dimensional nature of the manifold, characterizing its topology for interpolation is difficult. To overcome this problem, we apply a local linear embedding (LLE) method to map the high dimensional man-

ifold to a lower dimensional space, and perform the interpolation therein [79].

Mapping the manifold to a locally linear lower dimensional space enables us to locally approximate the topology of the manifold with the Delaunay triangulation [80]. The local structure of the manifold (in low dimension) is thus characterized by the d-simplex (e.g. triangle in $2D$ or tetrahedron $3D$), and we can now generate a sample on the manifold by interpolating it on the d-simplex, and then map it back to the high dimensional space: First, a d-simplex, along with its associated deformation field, is randomly selected from the manifold, and a d-dimensional random vector $\boldsymbol{r} \in \mathbb{R}^d$ is generated such that each of $\boldsymbol{r}$'s components are uniformly distributed on $[0, 1]$. $\boldsymbol{r}$ is then normalized so that $\|\boldsymbol{r}\|_1 = 1$. After that, a new deformation field can then be generated as $D^* = \sum_{i=1}^{d} r_i D_i$. Similar to the k parameter above, larger d results in more deformation fields contributing in the generation of the new one, which usually cause the new deformation to be smoother. The above process can be repeated indefinitely, and allows us to create arbitrary deformation fields.

It is noted that learning the object manifold and generating new objects are topics having been studied by many researchers [81, 77, 78, 82] and we are not claiming the proposed algorithm being superior to any of the existing ones. In fact, our key objective is to simulate a known deformation so that we can always keep track of the exact boundary of the nucleus, for the ultimate purpose of evaluating nuclear segmentation algorithms.

The final nucleus is determined as $D^* \circ P$. Moreover, the segmentation of the nucleus is also known, which is characterized by $D^* \circ Q$. This is the key that enables us in using such synthesized images for evaluating nucleus segmentation.

Local Structure Synthesis

The above nucleus synthesis addresses the appearence of each newly generated neucleus. It does not explicite its location. This is the subject of this section. Indeed, the spatially local arrangement of the nuclei is a indicator of local cell/tissue type. Epithelia cells are often tightly aligned whereas brain cells are mostly separated in an close to uniformly distributed fashion. Ev-

idently, one needs the category information of the nucleus to determine the closeness of its neighbors. This, however, requires another layer of work for manual labeling.

To alleviate such a problem, we take an approximated approach. First, slides from the same organ, e.g., lung, are grouped together. Within the group, a clustering is performed based on straightforward nuclear features including:

Area.

Circularity. The ratio between the two major axes

Intensity. The mean of hematoxylin component within the nucleus.

Speckleness. The standard variation of hematoxylin component in the nucleus.

The affinity propagation algorithm is used to cluster the nuclei into several groups.

For each group, a new location is randomly generated globally to place the first synthesized nucleus. Afterwards, the nuclei from the same group is only placed in a relatively local approach. Specifically, the Delauney triangulation is constructed among the centroids of the segmented nuclei. For each nucleus, the mean and standard deviation of distances to its neighbors are recorded. In generating a new neighboring nucleus, a random number is drawn from such a normal distribution and used as the distance to learn the distances to local neighbors.

In synthesis, first, only throw a few representative nuclei. Then, starting from them and grow their neihgbors according to the local distances to neighbors. This way we have epithelial cell very close to another epithelial cell.

Cytoplasm and Stroma

Once we have generated the nuclear regions, we can "fill in" the cytoplasmic and the inter-cellular regions. While the segmentation of the cytoplasm is

not available in our training data and in many times an ambiguous problem even for human eyes, it fortunately does not affect the present work significantly because we are mainly interested in the nuclei. In this work, we model the image content in the non-nuclear regions as a Markov Random Field (MRF) [83, 84].

More explicitly, we start with any pixel location $\boldsymbol{x}$ in the newly synthesized image U that is not in, but bordering, the nucleus. The choice of such a location is because with some nucleus structures around, it is easier to infer the image intensity at this location. Then, a neighborhood $\omega_{\boldsymbol{x}}$ is extracted centering at $\boldsymbol{x}$. According to the assumption, some pixels in $\omega_{\boldsymbol{x}}$ have already been known. Then, we find the most similar patch α in non-nuclear region of the training images. In the computation of similarity, the mean-square-difference (MSD) is computed over color image values of the already filled pixels. In addition, denote the similar patches as the set Γ:

$$\Gamma := \{\beta : MSD(\beta, U(\omega_{\boldsymbol{x}})) \leq 1.3 MSD(\alpha, U(\omega_{\boldsymbol{x}}))\} \tag{5.2}$$

Based on the MRF assumption, the distribution of the center pixels of all the patches in Γ is the same as that of the $U(\boldsymbol{x})$, weighted (inversely) by $MSD(\beta, U(\omega_{\boldsymbol{x}}))$. A random sample is therefore drawn from this distribution and assigned to $U(\boldsymbol{x})$.

After this, we move onto the next un-filled pixel and repeat the same procedure above, until all the pixels in the image have been filled.

5.2.2 Evaluation of Segmentation Algorithm

The proposed algorithm is able to synthesize arbitrarily large image data with known segmentation of the nuclei. With that, we can evaluate a given nucleus extraction algorithm at large scale.

Brief Description of Our Segmentation Algorithm to be Evaluated

Here, we briefly describe the nucleus segmentation algorithm that is to be evaluated. Given a new H&E stained digital pathology image $I : \mathbb{R}^2 \to C$, we want to extract the nuclei from it. To that end, we first normalize the

color distribution in the CIE Lab color space to correct for possible staining, imaging, and illumination artifacts. Then, the RGB image is separated into hematoxylin and eosin channels [10].

Then, in the hematoxylin channel, the Otsu threshold is computed to give the initial extraction of the nuclei. This is followed by the fine tuning of the local statistics driven level set evolution [85]. The resulting segmentation may consist regions where several nuclei are clumped together. In order to de-clump the region and obtain the definition of each individual nuclei, the mean shift algorithm is used [86, 87], which gives the final segmentation of individual nuclei.

The algorithm is implemented using the Insight Toolkit [88] and the OpenSlide library [89], which is able to run on large tiff-like images output from the microscopy scanner.

5.3 Results

5.3.1 Image Synthesis Results

Figure 5.1 demonstrates the synthesized brain pathology image results. Note that in addition to being visually realistic, the most desired feature is that the manual segmentation results have been fully transferred to the newly generated images, including "touching" nuclei that are also labeled correctly. Separating those nuclei is particularly important, and challenging for any nucleus segmentation algorithm.

5.3.2 Quantitative Evaluation Results

Performing nucleus segmentation in the above images gives the segmentation results as shown in Figure 5.2. It can be observed that in the left-most image, the two touching nuclei are correctly separated. While in the next image, the touching nuclei in the top-right corner are not separated correctly, forming a Mickey-mouse head shape. The two nuclei below Mickey-mouse are not detected as two nuclei at all. Such difficulty is expected when we generate the

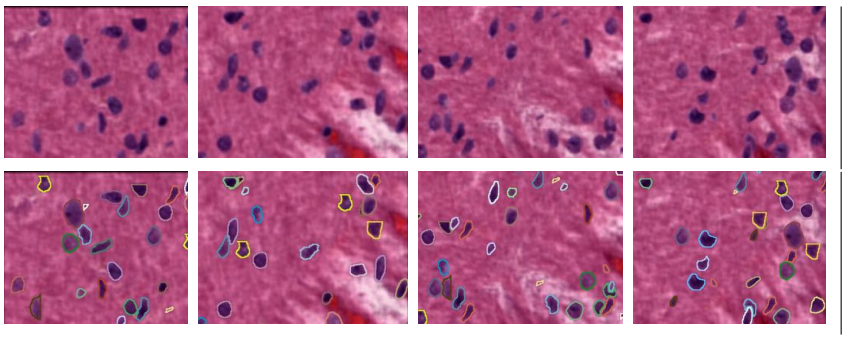


Figure 5.1: Synthesized images (the first row) with the ground truth segmentation (the second row). Note: the proposed synthesis algorithm naturally handles the labeling of the touching nuclei. This enables the evaluation of the segmentation algorithms ability in separating them, which is often time one of the most challenging step in the nucleus segmentation.

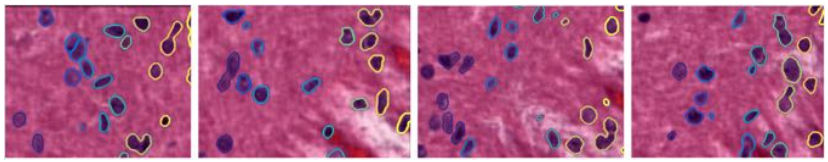


Figure 5.2: The segmentation algorithm run on the above synthesized images. In the second (from left) figure, the touching nuclei in the top-right corner are not separated correctly, forming a Mickey-mouse head shape. The two nuclei below Mickey-mouse are not detected as two nuclei at all.

images. More importantly, the purpose of the proposed synthesis framework is to enable *large scale* evaluation. To that end, 1000 images containing a total of half-million ground truth nuclei are synthesized. The segmentation algorithm is tested on those images and an average Dice coefficient of 0.71 is achieved with standard deviation of 0.04.

5.4 Conclusion and Discussion

In this work we proposed a method to synthesize arbitrarily large histopathology images from a small set of training images in which the nuclei have already been segmented. In doing so we can systematically generate large validated data sets to evaluate any given automatic nucleus extraction algorithms.

There are several limitations that will be topics of our future research. First, in the training image we do not have the cytoplasm information marked out. Therefore, in the synthesized images, while we do observe the texture information in the non-nuclear region, we do not have a clear definition for cytoplasm and cell boundary. This is acceptable for the current nucleus segmentation evaluation purpose. But this will need to be extended in a future more comprehensive study. Second, the global/topological information is not well presented. We present in the results section the images generated from brain slides. However, for tissues from other organs, meso-scale features such as crypts and glands are also very important. Since our synthesis algorithm works at a fine resolution only, the current framework does not generate such meso-scale structure. To perform multi-scale tissue synthesis is part of our on-going research.

We will further investigate on the synthesis in multiple scales: this will simultaneously solve the problems of synthesizing the very large WSI, as well as synthesizing structures at multiple size. Specifically, we will first synthesize, at low resolution, the entire whole slide image. In this image, on pixel spans roughly 8 micrometers. Then, we will dive into higher resolution but at more restricted tile size. In the synthesis at higher resolution, the already synthesized lower resolution pixel values will be used as prior in formation in searching for similar patches. By doing this, the image contents along multiple scales are naturally coherently connected. Moreover, spatially, the image contents across multiple tiles are smoothly patched.

Chapter 6

Evaluation of Nuclei Segmentation via Deep Image Synthesis

Multi-scale feature analysis renders the main theme for most image-genetics correlation study. Such joint computing of drastically heterogeneous features links information extracted from nano-to-millimeter, covering six degree of magnitudes. In particular, at sub-micron level, the giga-pixel digital histopathology images reveals much information of the cells and nuclei. Among the multiple observable features spanning several scales in the pathology images, nuclear morphology is one of the central criteria for diagnosis and grading. As a result, it is also the most studied target in computer aided analysis. Many research papers have been devoted to the problem of extracting nuclei from digital pathology images, which lies at the foundation of any further correlation study. The quality of the nucleus extraction, have only been evaluated at the level of about thousands of nuclei, whereas a single slide may contain almost a million. The lack of large scale systematic validation and evaluation poses a rather serious challenge especially when interpreting the results of the downstream correlative study. In this report, we aim to address such an issue through a deep neural network framework. Specifically, large scale images with ground truth nucleus morphology are synthesized and the sensitivity and specificity of several nuclear segmentation algorithms been evaluated. We believe such a quantitative evaluation framework could facilitate a more validated study for current and future histopathology images analysis field.

6.1 Introduction

Genome is not able to tell the entire trajectory of a human, due to the abscent of epigenetic and environmental influence in all the years after birth. Such information can be incorporated through a longitudinal homogeneous fashion, and it can also be captured from multiscale heterogenous sources. Image-genetics correlation study is one attempt in the later direction[90, 91].

Multiscale feature analysis renders the main theme for most image-genetics correlation study. Such joint computing of drastically heterogeneous features links information extracted from nano-to-millimeter, covering six degree of magnitudes. In particular, at sub-micron level, the gigapixel digital histopathology images (whole slide images, WSI) reveals much information of the cells and nuclei. Digital histopathology image computing is becoming an ever more active field due to its superb spatial resolution and the availability of large data sets. At a spatial scale between genomics and radiology, histopathology images have shown the promising possibility of revealing detailed image based features that can be linked with genetics in a straightforward manner, as well as the epigenetic and environmental factors, which are not present in the DNA sequences, and not visible in radiology imagery.

Among the multiple observable features spanning several scales in the pathology images, nuclear morphology is one of the central criteria for diagnosis and grading. As a result, it is also the most studied target in computer aided analysis. Many research papers have been devoted to the problem of extracting nuclei from digital pathology images, which lies at the foundation of many further correlation study [3].

There exists an extensive literature discussing nucleus segmentation [61, 62, 63, 64, 65, 66, 15, 67, 68, 69, 70]. The quantitative validation and evaluation of a given nucleus segmentation algorithm is one of the most challenging issues in all of the aforementioned publication. The quality of the nucleus extraction, have only been evaluated at the level of about thousands of nuclei, whereas a single slide may contain almost a million. Indeed, it is very difficult, if not impossible, to ask a skilled pathologist to manually contour millions of nuclei to be compared with algorithm output, which is the standard validation approach used in radiology images.

The lack of large scale systematic validation and evaluation poses a rather serious challenge especially when interpreting the results of the downstream correlative study. In this report, instead of relying on human annotation, we need a systematic approach to computationally generate large enough data set to test the accuracy and robustness of the algorithm. We propose a method to address such an issue through a deep neural network framework. Specifically, large scale images with ground truth nucleus morphology are synthesized and the sensitivity and specificity of several nuclear segmentation algorithms been evaluated. To the best of our knowledge, this work is the first addressing this problem at a large scale. We believe that this type of analysis is not only informative for algorithm developers to improve their techniques but also necessary knowledge to be provided to the end-user community whose analysis is based on nuclear segmentation algorithms.

6.2 Method

A segmentation evaluation platform has to provide two interconnected major components: first, image: we need to have a large set of images. Second, contours: one has to have the ground truth for each target, in the current case nuclei. In this section we first briefly review some texture synthesis methods and then present our proposed evaluation framework.

6.2.1 A Brief Review of Texture Synthesis Methods

The image/texture synthesis problem has been the main topic in many previous studies [84, 92, 93, 94, 95, 96, 97]. The target of texture synthesis is to learn a generative model from a small region of texture. When samples drawn from the model, they look like extensions of the same texture. Most synthesis algorithms can be categorized into two classes: exemplar based and feature model based. Exemplar based methods start with one or few points and look for the most similar patches from the original image. Once the similar patches have been identified, the corresponding point is picked from one randomly chosen patch. Such process is repeated at the pixel level or small block level until the entire new image is filled.

Contrastingly, the second class of method synthesizes a new image through matching the features in the transformed space. Conceptually, this is because

the texture synthesis is a problem dual to the texture classification. The texture classification constructs a feature space for certain type of texture and classifies all images with similar texture to be of the same type. From the other direction, one can generate a new image from the learned feature model, which by definition is of the same texture category.

In [97], the author demonstrates the generative power of their model while trained in a purely discriminative fashion. This approach explicitly defines a parametric texture model by the correlations between feature maps in some layers of the CNN. For notification, a layer with N_l distinct filters has N_l feature maps each of size M_l vectorized, which can be stored in a matrix $F^l \in R^{N_l \cdot M_l}$. The feature correlations are given by the Gram matrix $G^l \in R^{N_l \cdot N_l}$, where G^l_{ij} is the inner product between feature map i and j in layer l:

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk} \tag{6.1}$$

The feature correlations G^l_{ij} are spatial summary statistics on these feature responses to characterize source image. To generate a new texture that is similar to a given image, the author uses gradient descent from a white noise image to determine the result which matches the Gram matrix representation of the given image. The model is the optimization problem: minimizing the mean-squared distance between Gram matrix from original image and image to be generated:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G^l_{ij} - \hat{G}^l_{ij})^2 \tag{6.2}$$

$$L(\vec{x}, \hat{\vec{x}}) = \sum_{l=0}^{L} w_l E_l \tag{6.3}$$

$$\frac{\partial E_l}{\partial \hat{F}^l_{ij}} = \frac{1}{N_l^2 M_l^2} ((\hat{F}^l)(G^l - \hat{G}^l))_{ji}, \ if \ \hat{F}^l_{ij} > 0 \tag{6.4}$$

where $\vec{x}$ and $\hat{\vec{x}}$ are the vectorized original and generated images; $L(\vec{x}, \hat{\vec{x}})$ is the total loss; w_l are weighting factors of each layer to the total loss. The gradient $\frac{\partial L}{\partial \hat{\vec{x}}}$ can be readily computed using standard error back-propagation.

Due to the fact that the feature map based methods capture both local and global information with various levels of rotational, scaling, and translational invariances, they are considered to be able to synthesize more realistically looking images [93, 97]. As a result, our image synthesize and evaluation framework will be based on this group of methods. In particular, due to the outstanding performance of the deep convolutional neural networks (CNN), we will adopt such an architecture for our system.

6.2.2 Image Synthesis and Quantitative Evaluation Framework

The image synthesis starts with a set of training images $I_i : \Omega \subset \mathbb{R}^2 \to C; i = 1, \ldots, N$ where C denotes the RGB color space. For each I_i, the pathologist validated segmentation of the nuclei are denoted as label images $J_i : \Omega \to \{0\} \cup \mathbb{Z}^+$. In the label images, 0 denotes background (non-nucleus) and different integer values indicate different nuclei. Figure 6.1 presents two real histology images of lung adenocarcinoma, with their respective segmentations shown in yellow contours.

Texture Image Synthesis

Comparing Figure 6.1(a) and (c), we can observe that different image regions have very distinct image characteristics. The proposed synthesis algorithm should take such spatial variation into consideration. As a result, a deep neural network architecture is designed as shown in Figure 6.2.

As an example, Figure 6.3 and Figure 6.4(the right column) shows the images synthesized for various types of tissue. It can be seen that the synthesized images look visually realistic.

However, the major challenge in constructing the image segmentation evaluation system lies in the fact that one has to not only synthesize the image itself, but also, we have to provide the segmentation ground truth in the synthesized image. In particular, when the nucleus regions are touching each other, as shown in the expert manual segmentation in Figure 6.1 (b) and (d). This is particularly difficult in the CNN based methods due to the fact that newly generated image does not have a pixel-wise correspondence with the original one. How to synthesize a image using CNN and passing the

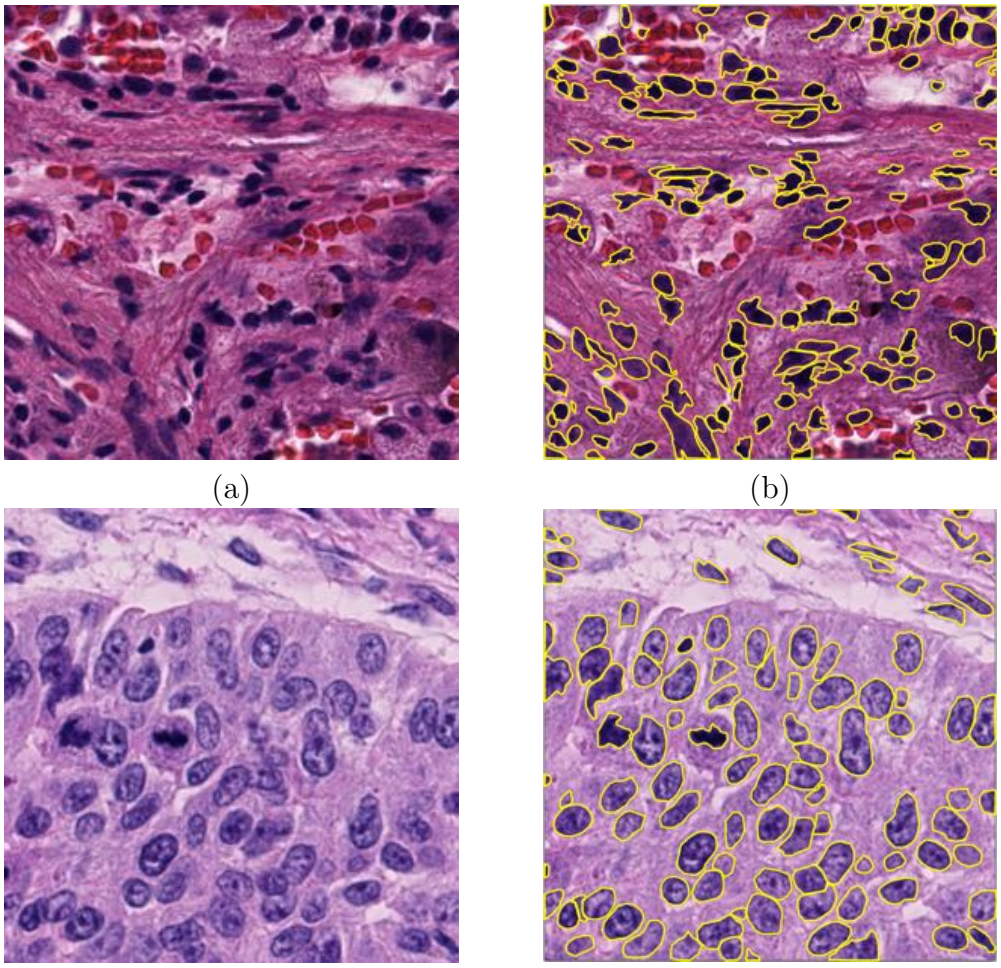


Figure 6.1: (a) (c) are two example patches from two WSI with the ground truth nuclei segmentations shown as contours in (b) (d). Note the two images have distinct appearance and the nuclei regions are "touching" each other

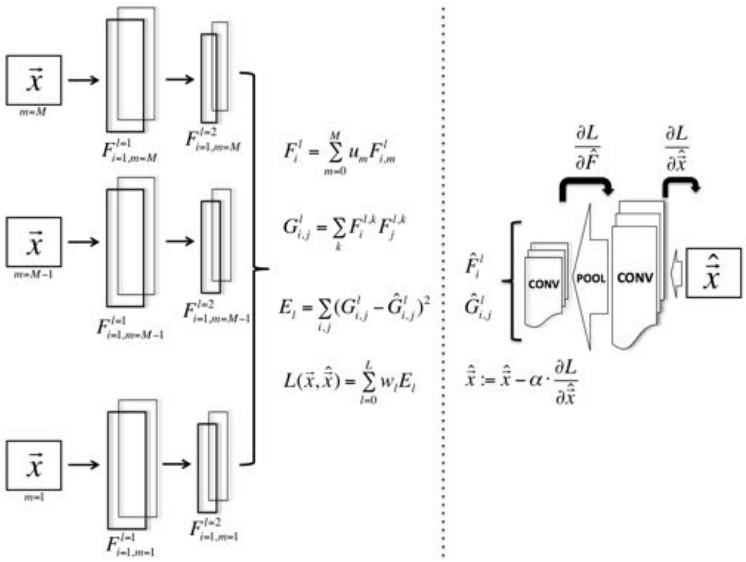


Figure 6.2: Deep neural network structure for image synthesis.

segmentation ground truth is the main challenge we address in the following discussion.

In subsequent discussion, this synthesize process, given any input image set $\mathscr{I}$ to generate a new image I', will be denoted as: $I' = \mathscr{S}(\mathscr{I})$.

The Characteristics of Image Synthesis with a Nuclear Mask

As having been eluded above, for each point $\boldsymbol{x}$ in the newly generated image $P : \Omega \to C$, it is not evident that whether $\boldsymbol{x}$ is within a certain nucleus. This is due to the inherent property of DNN based texture synthesize.

In order to synthesize an image and passing the label information along, we first investigate the fine details in the synthesized images. In particular, we will investigate how a region with constant color value will appear in the synthesized image. This will provide the insight for passing the nuclear label information.

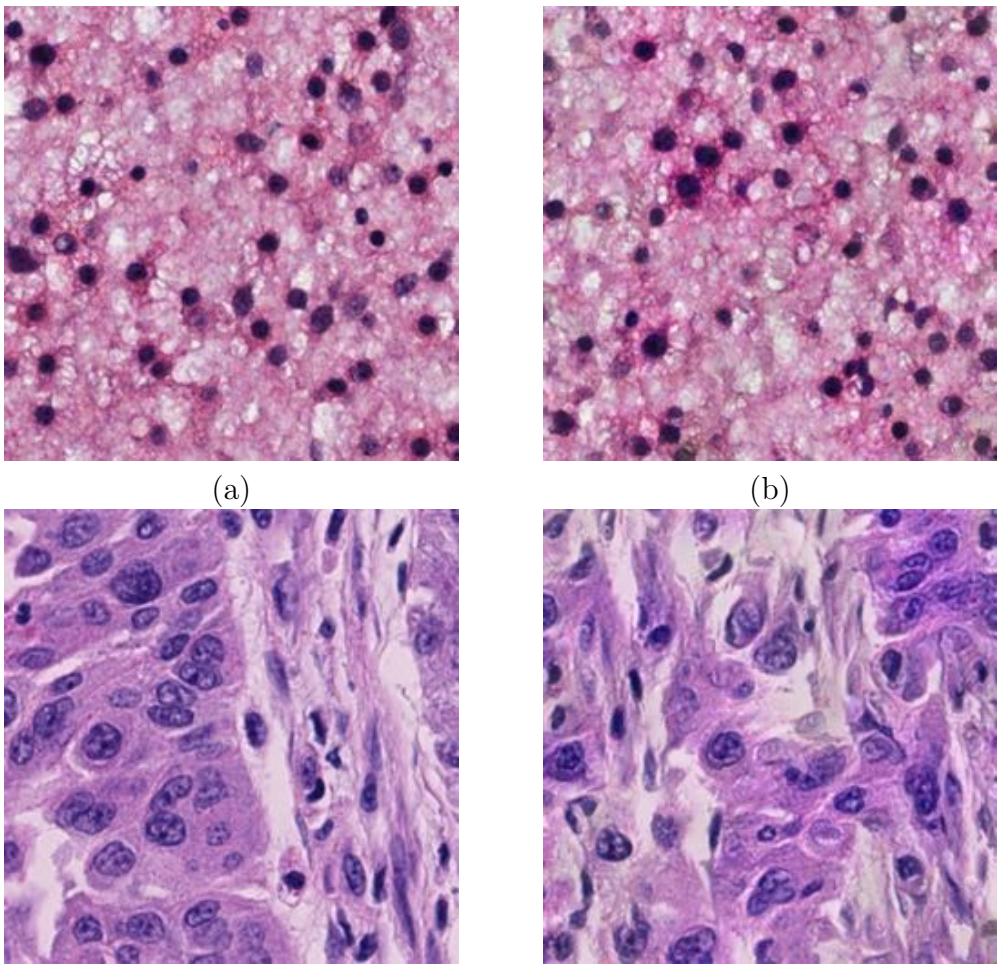
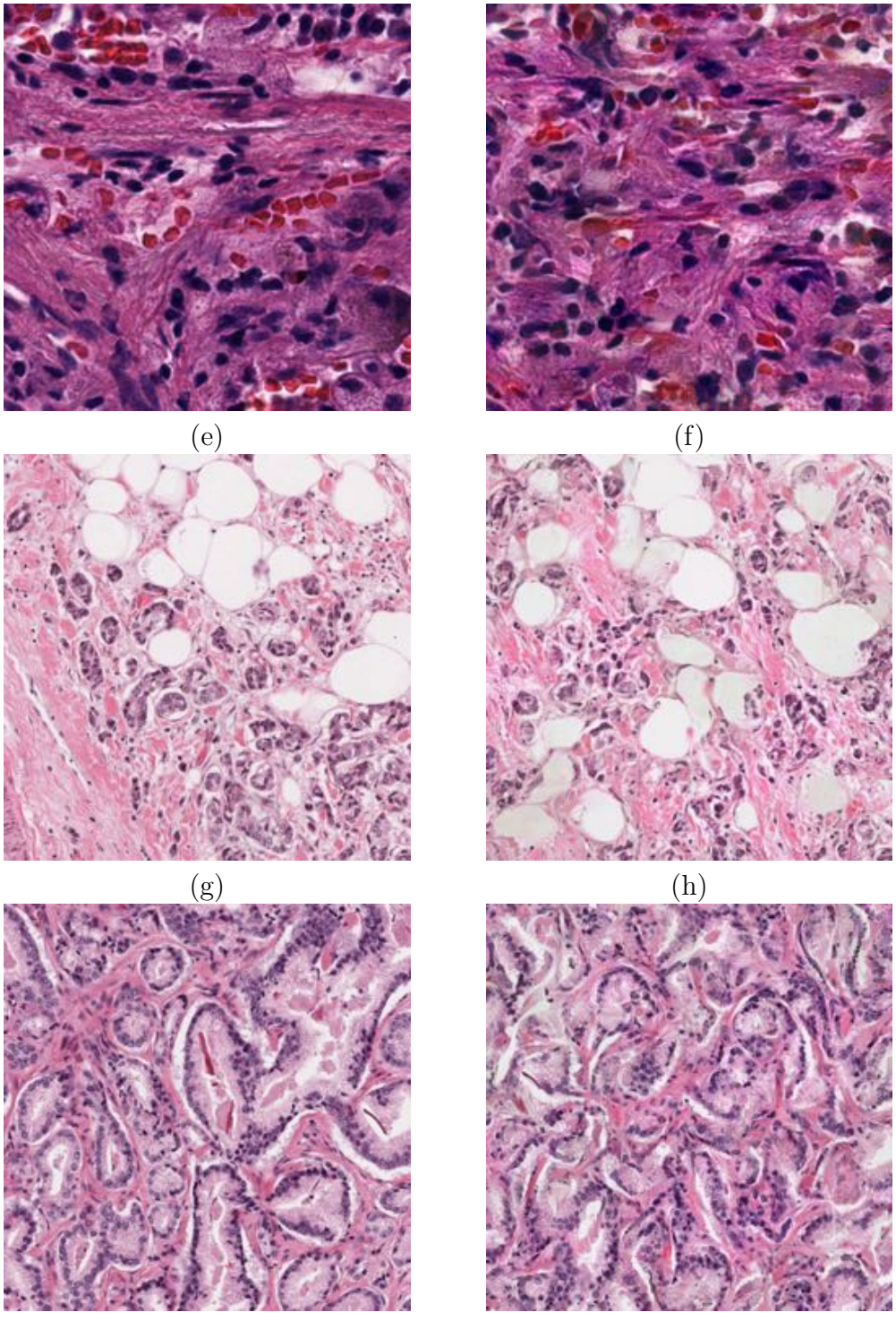


Figure 6.3: Image synthesized (b, d) for Lower Grade Glioma(LGG), Lung Squamous Cell Carcinoma(LUSC)(a, c respectively).



(e) (f)

(g) (h)

(i) (j)

Figure 6.4: Image synthesized (f, h, j) for Lung Adenocarcinoma(LUAD), Breast Cancer(BRCA), Prostate Adenocarcinoma(PRAD) (e, g, i, respectively).

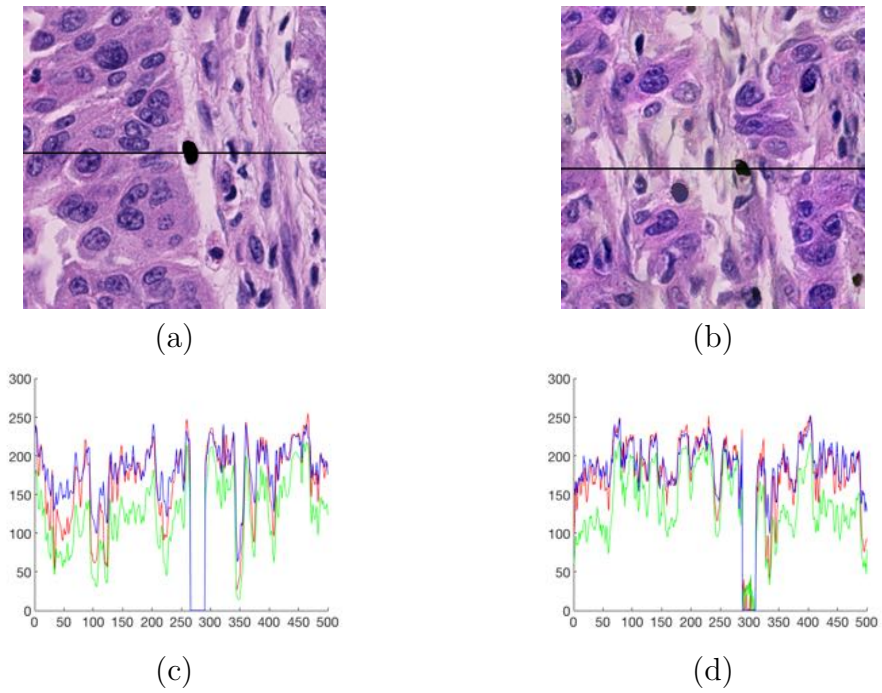


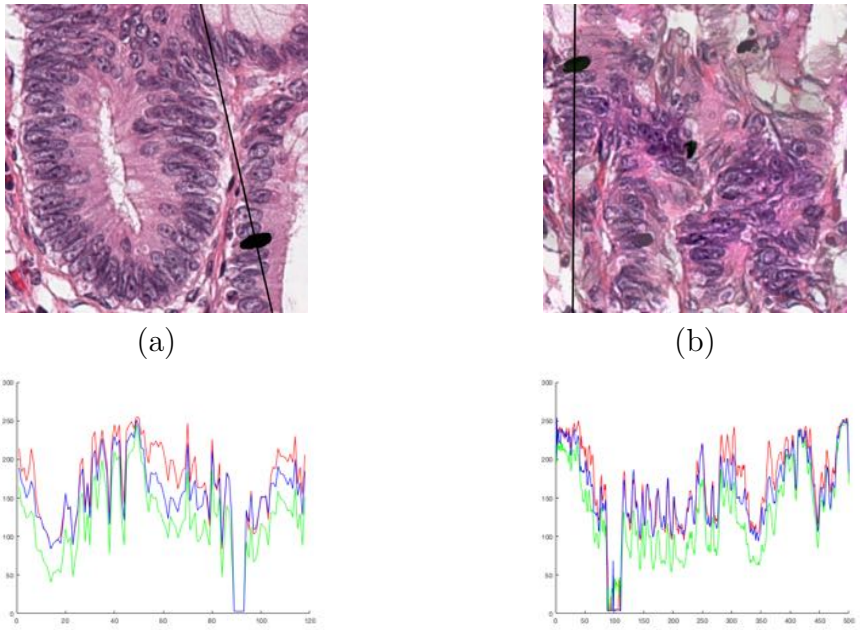
Figure 6.5: One case study where nucleus is not touching.

For simplicity, we start from only one training image I_1. In it, denote the regions D_1^i as $D_1^i := \{\boldsymbol{x} : J_1(\boldsymbol{x}) = i\}$. Then, with out loss of generality, let

$$K_1(\boldsymbol{x}) = \begin{cases} [0,0,0]^\perp & \text{if } \boldsymbol{x} \in D_1 \\ I_1(\boldsymbol{x}) & \text{otherwise} \end{cases} \tag{6.5}$$

That is, we paint the region of the first nucleus as pure black color. Using the method in Section 6.2.2, a new image $\tilde{K}_1$ is synthesized as $\tilde{K}_1 = \mathscr{S}(K_1)$. The images K_1 and $\tilde{K}_1$ are both shown in Figure 6.5(a and b, respectively).

From the color value profile we observe that: in the original image Figure 6.5(c), the colors in the nucleus are to zero. However, in the synthesized Figure 6.5(d), the distinction between its surrounding is much less evident. This is an expected effect due to the fact that, in the synthesis algorithm, the optimization goal is to matching the texture descriptor computed from the feature layer responses, not to image color. In theory, there is even no guarantee that the synthesized image would contain such a "dark spot". As long as the feature maps match, the synthesize process will converge. Fortunately,



(a) (b)

(c) (d)

Figure 6.6: One case study where nucleus are touching.

in the case of Figure 6.5, the nucleus is not in contact with any others. In such a case, one may try to set a threshold in the synthesized image to define the region of the nucleus. However, in the more common cases as shown in Figure 6.6, the nuclei are touching each other and to differentiate the specific nucleus in the synthesized image is particularly challenging. In the original image with one nucleus marked pure black as shown in Figure 6.6(a), the nucleus is easy to be distinguish from its neighbors, see also the single row profile Figure 6.6(c). However, in the synthesized image, the color values have a long transition zone and the exact nuclear boundary is very difficult to define, as shown in Figure 6.6(d).

Based on the experiments in Figures 6.5 and 6.6, two observations can be made. *First*, filling the nuclear region with a distinct color and then synthesize the new image using the filled image, it is able to define the nucleus in the new image. More generally, if all the nuclear regions are filled with a single color, then, in the synthesized image, the entire nuclear region can also be identified. *Second*, such a scheme is not able to specifically identify

different nuclei when they are spatially close to each other.

Because the ultimate goal of the image synthesis procedure is to evaluate the sensitivity and specificity of various nuclear segmentation algorithms. Inspired by the two observations above, we proposed an approach that evaluates the sensitivity and specificity of nuclear segmentation algorithms separately.

Image Synthesis for Segmentation Sensitivity Evaluation

Since the synthesize process does not transfer the label information, we have to "burn" the nuclear label information into the image and "decode" them after new image is synthesized. As we mentioned above, a region of constant color value will result in a region with varying colors in the synthesized image. Therefore, in order to minimize the color varying effect on the determination of the nucleus boundaries, one need to find the color that is "most distinct" from the existing color in the image. As a result, even with the color mixing effect, the decision on nucleus boundaries would be of less ambiguity.

To that end, an optimization problem is set up as:

$$c^* = \underset{c\in\mathbb{R}^3}{\operatorname{argmax}}\, E(c) \tag{6.6}$$

$$\text{s.t. } [0,0,0]^\perp \leq c \leq [255,255,255]^\perp. \tag{6.7}$$

In that,

$$E(c) := \frac{1}{|\Omega|N} \sum_{i=1}^{N} \int_{\boldsymbol{x}\in\Omega} (c - I_i(\boldsymbol{x}))^2 \, d\boldsymbol{x} \tag{6.8}$$

where $|\Omega|$ is the domain size.

Solving such a constraint optimization problem results in a color c^* which is furthest from the existing color in the training images and least likely to be mixed with existing tissue colors after synthesis.

After the "optimal filling color" c^* is identified, the entire nuclear region in the training images are filled with c^*:

$$P_i(\boldsymbol{x}) = \begin{cases} c^* & \text{if } J_i(\boldsymbol{x}) > 0 \\ I_i(\boldsymbol{x}) & \text{otherwise} \end{cases} \tag{6.9}$$

To synthesize a new image $\tilde{Q} = \mathscr{S}(\{P_i\}_1^N)$, the nuclear region in $\tilde{Q}$ is determined as $D_{\tilde{Q}} := \{\boldsymbol{x} \in \Omega : |\tilde{Q}(\boldsymbol{x}) - c^*| \leq \alpha\}$. Evidently in $\tilde{Q}$, the image is filled with color similar to c^*, which need to be in-painted. However, first, nuclei have rich internal textures. Second, the rich nuclear texture does not exist in the region outside $D_{\tilde{Q}}$. Therefore, regular in-painting algorithms which restore the missing values from surroundings do not meet this need and we need to seek other approaches. In particular, we would utilize the nuclear region in the training images to restore those in the synthesized image.

To that end, denote the entire nuclear region for each training image I_i as $B_i := \{\boldsymbol{x} \in \Omega : J_i(\boldsymbol{x}) > 0\}$. The similarity between $D_{\tilde{Q}}$ and each B_i is compared. Then, the nuclear texture in the most similar image is transferred to fill in the new image.

More explicitely, the distance E_i between $D_{\tilde{Q}}$ and each B_i is defined follows [73, 74]: A set of points $G := \{\boldsymbol{x}_i \in D_{\tilde{Q}} : i = 1, \ldots, m\}$ is sampled uniformly from $D_{\tilde{Q}}$, and a same number of points $H := \{\boldsymbol{y}_i \in B_i : i = 1, \ldots, m\}$ are sampled from each nuclear region. Each point is considered to have a Dirac metric. Then, an optimal matching between the two sets of points is constructed. To that end, we denote the correspondence between X and Y by a matrix $A \in \{0, 1\}^{m \times m}$ where $A_{i,j} = 1(0)$ indicates $\boldsymbol{x}_i$ is corresponding (not corresponding, resp.) with $\boldsymbol{y}_j$. Denoting the pair-wise distance matrix $F \in \mathbb{R}^{m \times m}$ as $F_{i,j} = \|\boldsymbol{x}_i - \boldsymbol{y}_j\|_2$ where $\|\cdot\|_2$ is the L_2 norm, we find the correspondence between the two sets of points by solving such an assignment problem:

$$\begin{aligned} A = \arg\min_{\tilde{A} \in \mathbb{R}^{m \times m}} & \|F \circ \tilde{A}\|_F \\ \text{s.t.} \quad \sum_j \tilde{A}_{i,j} &= 1 \quad \forall i \in \{1, \ldots, m\} \\ \sum_i \tilde{A}_{i,j} &= 1 \quad \forall j \in \{1, \ldots, m\} \\ \tilde{A}_{i,j} &\geq 0 \quad \forall i, j \in \{1, \ldots, m\} \end{aligned} \tag{6.10}$$

where $\circ$ is the Hadamard product of the two matrices and $\|\cdot\|_F$ is the matrix Frobenius norm. Moreover, it is noted that the optimization variable $\tilde{A}$ is not restricted to be a binary matrix. Otherwise the optimization becomes

an NP-hard combinatorial problem. On the other hand, due to the fact that the constraint matrix of (6.10) is totally uni-modular, the resulting optimal A is a binary matrix [75]. This optimization problem can be shown to be convex, and it can be effectively solved by using, for example, the interior point method [76]. The resulting matrix A will give a one-to-one correspondence between G and H. The deformation field D_i is therefore constructed as the displacement vector field among the corresponding points. Finally, the distance E_i between $D_{\tilde{Q}}$ and B_i is the optimal $\|F \circ \tilde{A}\|_F$.

Such an optimal matching construct a point-wise correspondence between $D_{\tilde{Q}}$ and the optimal B_i, and the image values of $Q(\boldsymbol{x})$ in $D_{\tilde{Q}}$ is then filled with those in B_i, following the mapping computed in A. As a result, we have a realistic looking new image $Q(\boldsymbol{x})$ in which the entire nuclear region is marked.

This image will be served as the ground truth for a nuclear segmentation algorithm. Furthermore, the texture synthesize procedure $\mathscr{S}(\cdot)$ can be executed with different initial random state, and produces different resulting images. As a result, such a process can be repeated arbitrarily many time, this is the key factor that enables the large scale evaluation of the sensitivity of any nuclear segmentation algorithm.

It is noted that such a synthesized image together with the label image of the entire nuclear region, is only able to evaluate the sensitivity of nuclear segmentation algorithm in extracting nuclear material. It is, however, not able to evaluate its performance on individual nucleus and/or separating touching nuclei. This is the subject in the next section.

Image Synthesis for Segmentation Specificity Evaluation

For solving the issue mentioned in the end the previous section, the labeling information for each nucleus, especially the touching ones, has to be synthesized. Unfortunately, as have been discussed in Section 6.2.2, the strong color variation makes it impossible to distinguish between adjacent nuclei.

In order to address this issue, in this section we take another approach. Instead of burn the label mask into the original image, here we burn the nucleus contours into the original image. Specifically, we denote the boundary of all nuclei in I_i, shown in Figure 6.7(a), as $T_i = \{\boldsymbol{x} \in \Omega : |\nabla J_i(\boldsymbol{x})| > 0\}$,

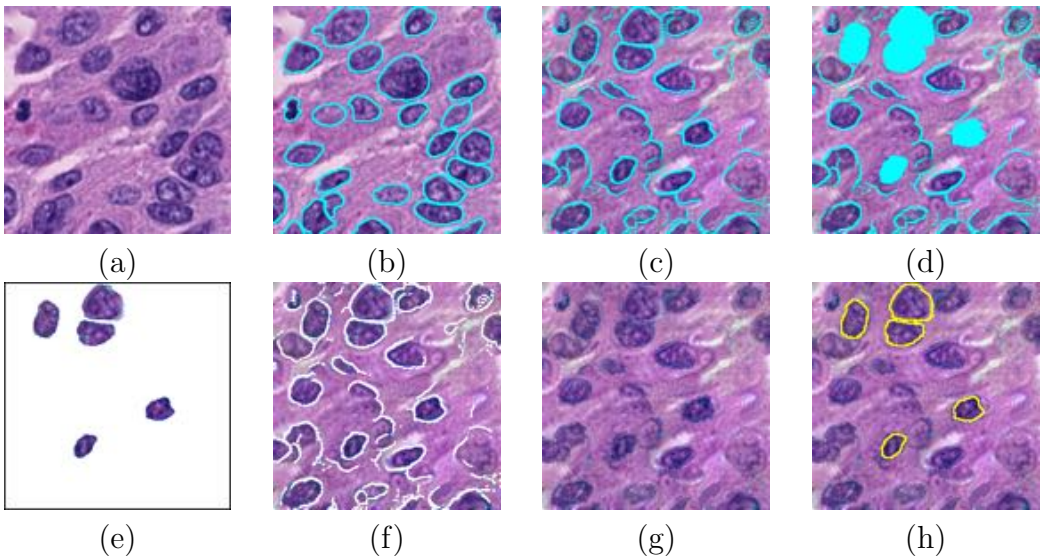


Figure 6.7: Synthesis through contour. (a)Original image; (b)Ground truth of image(a); (c)New image synthesized from image(b); (d)Filled nuclei with closed boundaries; (e)Subtraction area (original RGB value) of image(c) from image(d); (f)Image(c) with cyan color boundaries removed; (g)Image for segmentation: image(f) with blank boundaries in-painted; (h)Ground truth for segmentation evaluation: boundaries of image(e) overlaid on image(g).

shown in Figure 6.7(b). It can be seen that due to the discrete computation of $\nabla J_i(\boldsymbol{x})$, T_i is usually two-pixel thick where there is a one-pixel layer outside the nuclear regions. This property will be useful in the subsequent processing. Then, we define $U_i : \Omega \to C$ as

$$U_i(\boldsymbol{x}) = \begin{cases} c^* & \text{if } \boldsymbol{x} \in T_i \\ I_i(\boldsymbol{x}) & \text{otherwise} \end{cases} \tag{6.11}$$

Then, synthesize a new image $\tilde{U} = \mathscr{S}(\{R_i\}_1^N)$. In $\tilde{U}$, shown in Figure 6.7(c), the nuclear boundary region is determined as $T_{\tilde{U}} := \{\boldsymbol{x} \in \Omega : |\tilde{U}(\boldsymbol{x}) - c^*| \leq \alpha\}$. Two issues remains to be solved before $\tilde{U}$ can be used for segmentation evaluation.

First, the synthesis algorithm $\mathscr{S}(\cdot)$ cannot guarantee a closed contour in the original image results in another closed contour. We can observe that, while many contours are enclosing correct nuclear region, some are spurious

branches not forming closed contours. One possible strategy would be to close those branches. This, however, would create false positives that enclose non-nuclear regions. On the other hand, with the sensitivity examination in the previous section, it is not necessary here to label all the nuclei. Therefore, we only pick those nuclei that are correctly enclosed in the synthesis process.

To that end, first, define $\tilde{V}$ to be the characteristic function of the set $T_{\tilde{U}}$. Then, for any point $p \in \Omega$ such that $p \notin T_{\tilde{U}}$, if p can be connected to $\partial\Omega$ without passing the set $T_{\tilde{U}}$, we set $\mathring{V}(p) = 1$. For all other points $\boldsymbol{x} \in \Omega$, set $\mathring{V}(\boldsymbol{x}) = 0$. This way, we mark all the pixels inside closed contours in the set $T_{\tilde{U}}$, as shown in Figure 6.7(d).

As a result, the image $\mathring{V}$ labels a portion of nuclei in the image $\tilde{U}$. Based on Figure 6.7(c) and Figure 6.7(d), we can get closed nuclei region by subtracting Figure 6.7(c) from Figure 6.7(d). The subtraction result would be a binary mask of closed nuclei. We multiply this binary mask with Figure 6.7(c), then we can get closed nuclei regions, as shown in Figure 6.7(e).

Yet, the colors in $T_{\tilde{U}}$ need to be filled. However, comparing the restoring of the nuclear texture, first, here the set $T_{\tilde{U}}$ (as shown in the white part of Figure 6.7(f)) only occupies the boundary region which has a relatively small area. Second, the texture to be restored "under the $T_{\tilde{U}}$" is of the same characteristics of those around them. As a result, a PDE based inpainting algorithm is adopted for this purpose [98]. The result is shown in Figure 6.7(g), which is the synthesis result being supplied to segmentation algorithms. Simply, we can overlay the boundary of the closed nuclei onto Figure 6.7(g), which generates the ground truth of the synthesis result, as shown in Figure 6.7(h).

Therefore, Figure 6.7(g) is one of the synthesized images originated from Figure 6.7(a). Figure 6.7(h) is the ground truth of Figure 6.7(g). Indeed, not all nuclei are enclosed in the synthesized image. That is, such a "ground truth" does not contain all the nuclei to be segmented. On the other hand, when evaluating a certain nucleus segmentation algorithm, it could extract more nuclei than those labeled here. As a result, the evaluation metric should be designed accordingly to reflect such a situation.

Quantification Metric for Image Realism Evaluation

For deep learning based synthesis model, we use the optimization cost function $L(\vec{x}, \hat{\vec{x}})$: weighted feature correlations among all layers from the deep learning model we used, between original image $\vec{x}$ and synthesized image $\hat{\vec{x}}$ to evaluate how realistic our synthesized images are.

$$L(\vec{x}, \hat{\vec{x}}) = \sum_{l=0}^{L} w_l E_l \tag{6.12}$$

where $E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - \hat{G}_{ij}^l)^2$, and $G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l$. The detailed explanations could be found in Section 6.2.1

Theoretically, this cost function quantifies the feature correlation between original image and synthesized images based on feature maps in deep learning models. The smaller this metric is, the more realistic our synthesized images are. This metric could also be applied to the exemplar based synthesis model we introduced in Chapter 5.

6.3 Evaluation of Nucleus Segmentation Algorithms

With the image synthesis framework setup, this enables the evaluation of nuclear segmentation algorithms at an arbitrarily large scale.

6.3.1 Separate metric for sensitivity and specificity

In the synthesis step, the different images are synthesized to test different aspects of the segmentation algorithms. In the evaluate step, we also need to measure the segmentation quality differently. Before that, we denote the two frequently used quantities for measuring the (dis-)similarity between the set G of the pixels annotated as a ground truth object and S a set of pixels segmented as a nucleus object. The Dice coefficient and the Hausdorff distance (HD) are defined, respeciively, as:

$$\Gamma(G, S) := \frac{2|G \cap S|}{|G| + |S|} \tag{6.13}$$

$$\Theta(G,S) := \max\{\sup_{x\in G}\inf_{y\in S}\|x-y\|, \sup_{y\in S}\inf_{x\in G}\|x-y\|\}. \tag{6.14}$$

First, we evaluate the sensitivity of the segmentation algorithm in correctly identify the nuclear region. To that end, the color image $Q(\boldsymbol{x})$ as well as the image $D_{\tilde{Q}}$ that masks the entire nuclear region are generated as in Section 6.2.2. $Q(\boldsymbol{x})$ is input to the nucleus segmentation algorithm $\mathscr{A}$, which compute a segmentation label image G. In some cases, G may be a labeled image where different nuclei are labeled with different values. However, in the evaluation of sensitivity, we treat all the labeled region as a single object, and the Dice coefficient and the HD are computed between $D_{\tilde{Q}}$ and G.

Second, the specificity of the segmentation algorithm, that is, the ability of that to articulate each individual nucleus, is evaluated. To that end, color image $\tilde{U}$ and the nucleus label image $\mathring{V}(\boldsymbol{x})$ are generated through the procedure in Section **??**. Applying the segmentation $\mathscr{A}(\tilde{U})$ to obtain the nuclear label image H. Here, it is likely that H may contain more regions than $\mathring{V}(\boldsymbol{x})$ due to the way it is constructed. However, we confine the evaluation only on those nuclei already marked in $\mathring{V}(\boldsymbol{x})$.

Inspired by the multiple object evaluation framework at [99], we denote: $\mathcal{G}$ as the set of ground truth objects in image $\mathring{V}$. $\mathcal{S}$ denote a set of algorithm segmented nuclei in image H. Furthermore, $\tilde{G}_i \in \mathcal{G}_a$ denote the i-th ground truth nucleus, and $\tilde{S}_i \in \mathcal{S}$ denote a segmented object that maximally overlaps $\tilde{G}_i$. n_S, n_G denote the total number of segmented objects in $\mathcal{S}$ and $\mathcal{G}$, respectively.

We define the object-Dice coefficient as:

$$\Gamma_o(G,S) := \sum_{i=1}^{n_G} w_i \Gamma(\tilde{G}_i, \tilde{S}_i) \tag{6.15}$$

and the object-HD is defined as:

$$\Theta_o(G,S) := \sum_{i=1}^{n_G} w_i \Theta(\tilde{G}_i, \tilde{S}_i) \tag{6.16}$$

where

$$w_i = |\tilde{G}_i| / \sum_{j=1}^{n_G} |\tilde{G}_j| \tag{6.17}$$

6.3.2 Brief Discussion of Segmentation Algorithm

Three nucleus segmentation algorithms will evaluated and they are briefly reviewed here.

Level-set based hierarchical mean shift method

Here, we briefly describe the nucleus segmentation algorithm that is to be evaluated. Given a new H&E stained digital pathology image $I : \mathbb{R}^2 \rightarrow C$, we want to extract the nuclei from it. To that end, we first normalize the color distribution in the CIE Lab color space to correct for possible staining, imaging, and illumination artifacts. Then, the RGB image is separated into hematoxylin and eosin channels [10].

Then, in the hematoxylin channel, the Otsu threshold is computed to give the initial extraction of the nuclei. This is followed by the fine tuning of the local statistics driven level set evolution [85]. The resulting segmentation may consist regions where several nuclei are clumped together. In order to de-clump the region and obtain the definition of each individual nuclei, the mean shift algorithm is used [86, 87], which gives the final segmentation of individual nuclei.

The algorithm is implemented using the Insight Toolkit [100] and the OpenSlide library [89], which is able to run on large tiff-like images output from the microscopy scanner.

Level-set based Watershed Method

Authors in [69] proposed a watershed based segmentation algorithm which is applied to the segmentation of whole slide images of the glioblastoma data in TCGA. The watershed algorithm is applied as another de-clumping approach. The previous steps in level-set based watershed are the same as those in level-set based hierarchical mean shift. We will evaluate this algorithm too.

Deep Learning Method

In Chapter 4, I proposed a deep learning model for nuclei segmentation of four types of cancer. It can separate clustered nuclei more accurately be-

cause I used hard negative edge mining, pixel-wise supervision and very deep CNN models. Of all of those CNN models, I selected the most effective one (30-layer Straight Model) to segment nuclei for this case in Section 6.4.2.

It is noted that we only tested a small subset of the nucleus segmentation methods, as very few are available and reimplementing all these algorithms is beyond the scope of this work.

6.4 Experiments and Results

In this section, we first present the image synthesis on Lung Squamous Cell Carcinoma (LUSC) tissue in Section 6.4.1. Then, in Section 6.4.2, the synthesized images will be used for quantitatively evaluate four different algorithms.

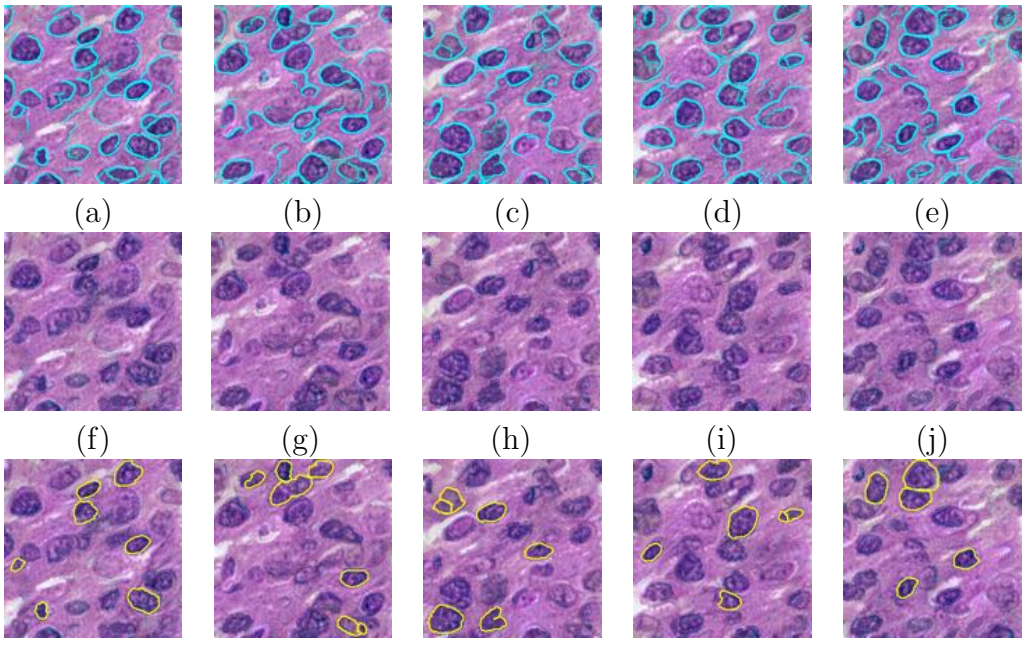


Figure 6.8: Five example images synthesized from Figure 6.7(b). (a)-(e)New images synthesized from Figure 6.7(b); (f)-(j)Images for segmentation; (k)-(o)Ground truth for segmentation evaluation.

6.4.1 Image Synthesis

We generate 5 new textures from Figure 6.7(b) using the methods described in 6.2.2, 4th Section: Image synthesis for segmentation specificity evaluation. The 5 new textures are displayed in Figure 6.8(a)-(e). As described in the 4th Section in 6.2.2, we generate in-painted images for segmentation algorithms, displayed in Figure 6.8(f)-(j). The corresponding ground truth of Figure 6.8(f)-(j) are displayed in Figure 6.8(k)-(o). In reality, we could generate arbitrary number of image and ground truth based on single image in this way. The four segmentation algorithms are going to be applied on Figure 6.8(f)-(j). Then, evaluation metrics will be computed by the segmentation results and the ground truth (Figure 6.8(k)-(o)).

6.4.2 Algorithm Evaluation

We qualitatively evaluate four segmentation algorithms based on the ground truth (shown in Figure 6.9(a)-(e) again for comparison) we generated in Section 6.4.1. The first algorithm no de-clumping method. It contains all the steps in the level-set based hierarchical mean shift method without de-clumping, as described in 6.3.2, 1st Section. The no de-clumping segmentation results are shown in Figure 6.9(f)-(j). The second algorithm is the level-set based hierarchical mean shift method in 6.3.2, 1st Section. The mean shift segmentation results are shown in Figure 6.9(k)-(o). The third algorithm is the level-set based watershed method in 6.3.2, 2nd Section. The watershed segmentation results are shown in Figure 6.9(p)-(t). The fourth algorithm is the deep learning method in 6.3.2, 3rd Section. The deep learning segmentation results are shown in Figure 6.9(u)-(y).

The quantitative segmentation results are shown in Table 6.1. The Table is separated into four parts corresponding to the four segmentation algorithms. We compute Dice 1, Dice 2 and Hausdorff Distance for each image. The Dice 1 is the evaluation metric described in Equation 6.12 and 6.14. While the Dice 2 is the similar metric conducted on each nucleus. The Dice 2 is described in Section 4.3.1. The Hausdorff Distance is described in Equation 6.13 and 6.15. We also compute the statistics for each evaluation metric of four algorithms. In terms of Dice 1, the deep learning method outperforms other three methods. But the Hausdorff Distance of the deep learning method is larger than those of the others. Based on our synthesis and segmentation

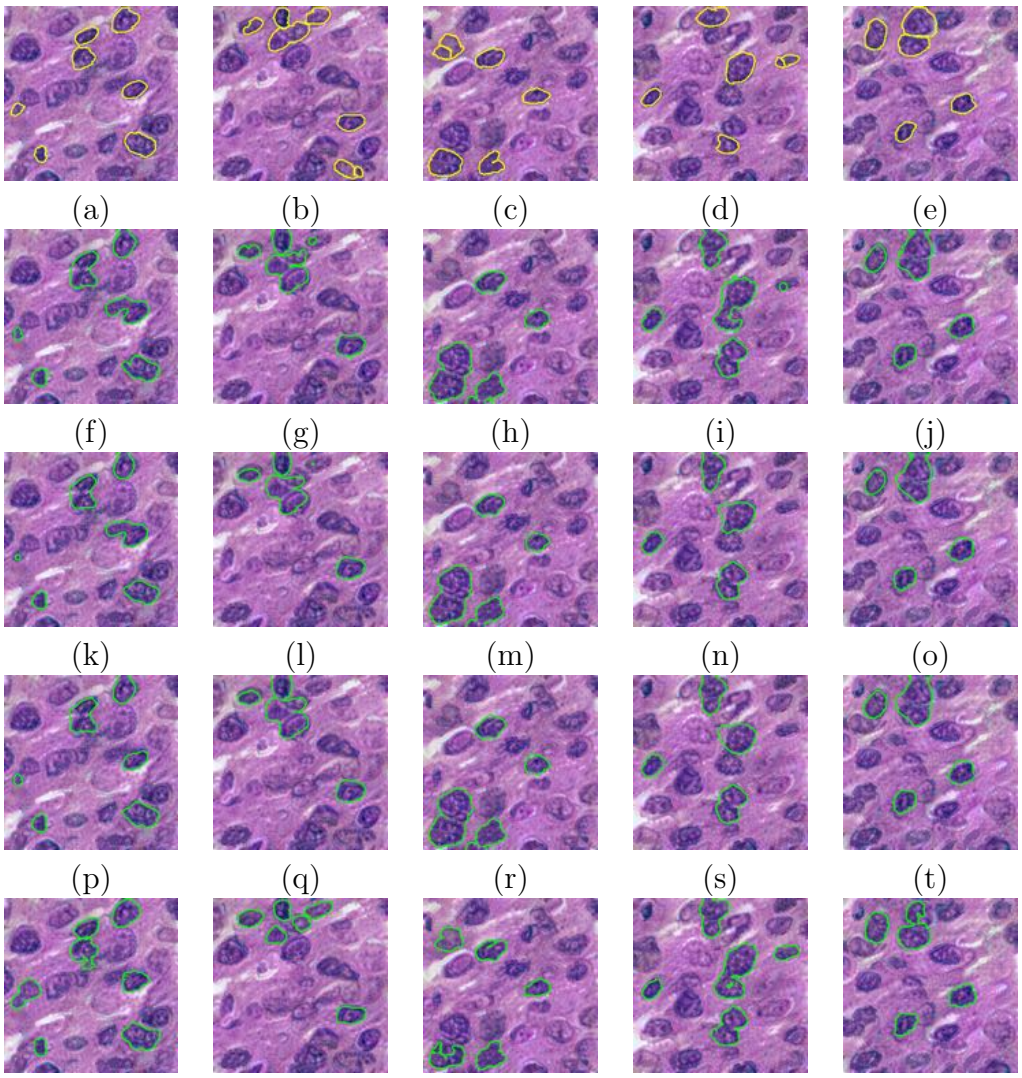


Figure 6.9: Four segmentation algorithms results of Figure 6.8(f)-(j). (a)-(e)Ground truth; (f)-(j)No-declumping method; (k)-(o)Mean shift method; (p)-(t)Watershed method; (u)-(y)Deep learning method.

Table 6.1: The segmentation evaluation results.

no declumping	**1**	**2**	**3**	**4**	**5**	**statistics**
Dice 1	0.746	0.553	0.631	0.595	0.827	**0.670±0.113**
Dice 2	0.234	0.193	0.284	0.242	0.332	**0.257±0.053**
Hausdorff Distance	4.899	5.745	4.796	5.831	5.657	**5.386±0.496**
mean shift	**1**	**2**	**3**	**4**	**5**	**statistics**
Dice 1	0.730	0.541	0.631	0.630	0.856	**0.678±0.120**
Dice 2	0.220	0.183	0.282	0.284	0.321	**0.258±0.055**
Hausdorff Distance	4.583	5.831	4.796	5.477	5.657	**5.269±0.549**
watershed	**1**	**2**	**3**	**4**	**5**	**statistics**
Dice 1	0.793	0.533	0.623	0.635	0.831	**0.683±0.125**
Dice 2	0.234	0.226	0.284	0.294	0.342	**0.276±0.048**
Hausdorff Distance	3.742	6.245	5.000	5.196	5.657	**5.168±0.930**
deep learning	**1**	**2**	**3**	**4**	**5**	**statistics**
Dice 1	0.791	0.708	0.818	0.647	0.796	**0.752±0.072**
Dice 2	0.231	0.193	0.275	0.277	0.315	**0.258±0.047**
Hausdorff Distance	5.477	5.916	5.196	5.745	4.690	**5.405±0.484**

evaluation framework, we can compare different segmentation algorithms by statistical significance computation. In general, it could be claimed that the large statistical evaluation is more objective than the small data set evaluation.

6.5 Conclusion and Discussion

In this work, we proposed a deep learning based method to synthesize arbitrarily large histopathology images from a small set of training images. The synthesis process is a generative image update using a deep discrimi-

native model. By this way, we can evaluate any given automatic nucleus extraction algorithms based on systematical generation of large validated data sets. In order to transfer the nuclei ground truth in the original small set of training images, we design two approaches: sensitivity evaluation and specificity evaluation. For the latter one, we provide 5 image synthesis results in Section 6.4.1. The four segmentation algorithms are applied on those 5 synthesized images. The qualitative and quantitative results are shown in Figure 6.9 and Table 6.1. Generally, we provide a feasible framework to solve the large ground truth demand.

There are several limitations which could be methods of our future projects. For example, we do not incorporate histopathological specific feature filters in the deep learning generation approach. Based on the principles described in Section 6.2.2, the parameters such as convolutional kernels are pre-trained for Gram matrix computing during texture generation. Therefore, the task is to design specific architecture of convolutional neural networks before parameter fine tuning. For example, histopathology image has object level features and spatial arrangement features. We will use different feature scales to design the convolutional kernels for different types of cancer. For the method to do texture similarity evaluation, we can compare the mean-squared distance between Gram matrix computed from different architectures, which is exactly the minimization item in this optimization problem. The smaller the mean-squared distance an architecture produces, the higher the texture similarity the architecture offers.

Chapter 7

Summary

There are five complete research projects in this dissertation: Intermediate Prostate Cancer Classification (Chapter 3); Nuclei Segmentation via Deep Learning (Chapter 4); Evaluation of Nuclei Segmentation via Image Synthesis (Chapter 5); Evaluation of Nuclei Segmentation via Deep Image Synthesis (Chapter 6); Clinical Information Visualization (Appendix A).

In Chapter 3, I proposes a new approach for intermediate prostate cancer grading, which makes large scale processing possible. I introduce deep learning architecture trained by optimized color decomposed image patches in localized areas. Our framework could achieve feasible classification accuracy.

In Chapter 4, I present a fully automatic end-to-end deep learning framework for nuclei segmentation for various types of cancer. Our method outperforms existing state-of-the-art nuclei segmentation methods. Our network uses no pooling nor stride convolutional layers, and enhances the edge representation capability of the deep convolutional networks by hard negative mining and pixel-wise supervision. This simple yet effective nuclei segmentation method is deployed without pre-processing, post-processing nor specific domain knowledge.

In Chapter 5, our nuclei segmentation evaluation framework could systematically generate large validated data sets to evaluate any given automatic nucleus extraction algorithms. This exemplar based image synthesis approach provides newly synthesized nuclei, cytoplasm and tissue separately,

which needs to be refined.

In Chapter 6, I propose a deep learning based method to synthesize arbitrarily large histopathology images from a small set of training images. The synthesis process is a generative image update using a deep discriminative model. It solves the nuclei tissue separation issue in Chapter 5, which is a fast and integrated scheme.

In Appendix A, I propose a practical framework for visual exploration of comorbidity between diseases. Experimental results show the goodness of SFDP graph drawing algorithm on large-scale data set, which also provide clinical meaning and nice visualization for future clinical oncology information research.

Appendices

Appendix A

Clinical Information Visualization

We present a practical framework for visual exploration of co-morbidities between diseases. By utilizing high-quality multilevel layout and clustering algorithms, we have implemented an innovative two-layer multiplex network of human diseases. Specifically, we extract the International Classification of Diseases, Ninth Revision (ICD9) codes from an Electronic Medical Records (EMRs) database to build our map of human diseases. In the lower layer, the abbreviated disease terms of ICD9 codes in the irregular regions look like cities in geographical maps. The connections represent the disease pairs co-morbidities, calculated by using co-occurrence. In the upper layer, we visualize multi-object profile of clinical information. For practical application, we propose an interactive system for users to define parameters of representations of the map (see a map representation example in Figure A.1). The demonstrated visualization method offer an opportunity to visually uncover the significant information in clinical data.

A.1 Introduction

EMRs contain insightful clinical information for assessing disease risk [101]. In the diagnosis table of an EMRs database, disease comorbidity refers to multiple diseases co-occurring in same patients more than chance alone [102]. Due to its implication for understanding human health, it is universally

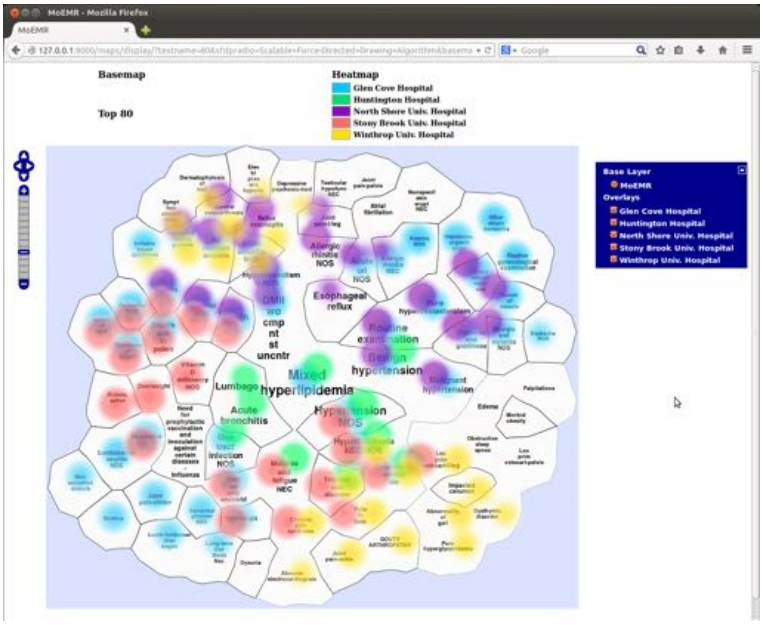


Figure A.1: A map of top 80 human diseases rendered in the web browser. The base map specification: Force Directed Placement (FDP) drawing algorithm, load factor: 5, edges disabled, coloring disabled. The heat map specification: the patient number density distributions of 5 large hospitals on Long Island.

regarded as an important healthcare research topic. Many statistics and bioinformatics research have been conducted on comorbidity to invent computational inference tools [103, 104]. In clinical informatics field, we hope to provide an efficient and effective visualization technique for EMRs data mining, especially comorbidity analysis. The main challenges for this interdisciplinary study lie in effective computation of diagnosis data, embedding comorbidity network (represented as graph strings) and clinical information highlight as a heat map overlay.

A recent work [105] on visual exploration of research papers in computer science field has addressed most of our visual analytics issues. It is an open-source visualization system based on graph drawing and map representation for data co-occurrence in large-scale relational database. Thanks to their open source github repository, we leveraged their document processing codes

on our EMRs data (Diagnosis Table) to calculate similarity matrix of disease pairs. The disease terms are represented as nodes and the connections of high comorbidity are represented as edges. We also use Graphviz [106] open source graph drawing library to visualize graphs in our approach. In the graph embedding and clustering part, we choose the scalable force directed algorithm and modularity based geometric clustering algorithm respectively. The two algorithms are highly related [107], because the spring model layout is consistent with clustering by optimal modularity. For clinical information overlay, we created multiple objects in our program to demonstrate categories in one type of clinical information (e.g. male and female in gender). At the same time, we manipulate the issue of heat spots overlap on single node by displacing heat spots in different directions by a small proportion. In the system, we nicely enable selections for top disease number, base map color representation, connection display and clinical information type.

In general, the main contributions of Maps of Human Disease are: First, we generate an esthetically symmetric layout of disease comorbidity by scalable force directed placement (SFDP), a graph drawing algorithm, which demonstrates relationships between prevalent diseases nicely. Second, differently from [105], we develop a multi-object heat map overlay solution to visualize multiple information profiles simultaneously on single base map. This is a very efficient mechanism for users to compare and discriminate clinical information via disease comorbidity within one visual space. Finally, we propose an interactive browser-based system for users to do parameter selection prior to the process of map generation.

A.2 Related Work

Analyzing and visualizing information in EMRs have been conducted to guide the diagnosis of future patients or to be used in studies of a certain disease. A number of researchers have explored visualization techniques for organizing the patient records in temporal event sequences [108], in which an outflow is proposed to summarize temporal event data that has been retrieved from EMRs. Other works focused on building an intuitive multi-faceted assessment of the patient overall information [109]. Interesting functionalities and hierarchies via body-centric data arrangement and standard health care coding support doctors or physician's review and analytical reasoning. Par-

ticularly, these EMRs representations typically operate without disease co-occurrence visualization, which enlightens us to apply the state-of-art graph-drawing methods on disease co-morbidity.

In the respect of research on human disease network, comprehensive approaches have been investigated on disease-related gene clusters detection [110], properties of the Disease Gene Network [111], and molecular relationships among distinct phenotypes [112]. Few of them discussed disease network layout algorithms in their research paper, except NodeXL, a free and open-source network analysis and visualization software package for Microsoft Excel. NodeXL provides well-known Force-directed graph drawing layout algorithms such as Fruchterman-Reingold [113] and Harel-Koren [114]. Another interesting work developed a phenotypic human disease network derived from two statistical measures quantifying associations between diseases [115]. In their two layer map, they used different colors to group nodes on the base map by first letter of the ICD 10 codes. The heat map shows significant group of related diseases over age and gender. However, their maps do not show any text information (eg. disease terms) on the nodes. Furthermore, they have not applied any geographic components to strengthen the map metaphor, which leads to poor vision effect.

A.3 Data and Methods

The raw data for this research is Practice Fusion De-Identified Data Set downloaded from https://www.kaggle.com/c/pf2012/data. We use records from two of the tables: the Diagnosis Table and the Patient Table. The Diagnosis Table consists of three columns: DiagnosisID (Primary Key), PatientID (Foreign Key) and ICD9Code. The Patient Table consists of three columns: PatientID (Primary Key), Gender and YearOfBirth. The total number of patient is 10,000. The hospital information in our system is fictitious.
The three main steps in our system is 1) generating comorbidity graph from comorbidity matrix, 2) generating base map via several graph related algorithms and finally, 3) generating clinical information heat map. Before the first step, we preprocess the raw data by merging the diagnoses (represented by ICD9 codes) belong to every patient together as an entity. In the process of map representation when disease terms needed, we create a look up table containing ICD9 codes and corresponding disease terms, which allow short

access time.

A.3.1 Comorbidity Graph Generation

In this thread, we defined two variable parameters, 'top number' as the number of nodes in the final comorbidity graph, and 'load factor' as the number of emitting edges from each node. The top number is used in step 3.1.1 and 3.1.2, and the load factor is used in step 3.1.3.

3.1.1 Disease Term Ranking:

Once the entities are loaded into the memory, we rank the ICD9 codes according to their occurrence frequency. Each code's weight is assigned by the times it occurred within the Diagnosis Table. The ranking list is acquired by sorting the codes weights and slicing based on the 'top number'. In our maps, the font size of each node is proportional to its rank in this list, clinically expressed as prevalence rate.

3.1.2 Comorbidity Matrix Computation:

Comorbidity matrix is the measurement to quantify edge lengths in the final comorbidity graph. Every element in comorbidity matrix is the pairwise similarity values between top diseases. We use Jaccard coefficient [116] to accommodate the Boolean nature of ICD9 codes in the entities. The Jaccard similarity coefficient is defined as the size of the intersection divided by the size of the union of the sample sets. In our study, pairwise comorbidity will be calculated within every pair of 'top' diseases:

$$J(S_i, S_j) = \frac{S_i \cap S_j}{S_i \cup S_j},$$

where S_i and S_j are the sets of patientID with disease i or j in the Diagnosis Table.

3.1.3 Edge Number Filtering and Edge Length Calculation:

Considering the visual clarity of the map, we select most highly related diseases from S_j for each disease in S_i ('load factor' = 1). In the next stage, the pairwise comorbidity matrix is transformed into a matrix of edge lengths for graph drawing. Since the value of $J(S_i, S_j)$ is in range of 0 and 1, in order to get an appropriate distribution, we need as well normalize $J(S_i, S_j)$,

by dividing $\overset{p=q}{max_{p,q}} J(S_p, S_q)$, thus we get rescaled value $J'(S_i, S_j)$. Based on $J'(S_i, S_j)$, each edge length of two nodes, $D(S_i, S_j)$ in the graph, is calculated by logarithm scaling:

$$D(S_i, S_j) = -E \cdot log[(1 - \sigma) \cdot J'(S_i, S_j) + \sigma],$$

where E is a scaling factor, we set it to be 1, and σ is a smoothing value, set to be 0.05. Logarithm scaling provides a better map metaphor by enlarging comorbidity values within the small range [105].

A.3.2 Base Map Generation

Base map generation is the most important thread in our approach. We apply Scalable Force Directed Placement (SFDP) [117], a fast graph drawing algorithm that efficiently layout large graphs, on large map generation. In the following part, we use a modularity based clustering algorithm [118] to group vertices into clusters. Lastly, in order to uncover beautifully the underlying structural information and neighborhoods, we apply the mapping and the coloring algorithms described in GMap [119].

3.2.1 Embedding:

In our system, we generate the positions of the nodes in two-dimensional plane by a spring-embedder [120], which relies on spring forces and mechanical energy based on Hooke's law. Basically, in the spring model, repulsive forces between all nodes and attractive forces between directly connected nodes are calculated iteratively to assign an energy state for each node, thus achieving a minimal energy state by moving each node based on the resultant force. In this basic model, the geometric layout is determined by the initial state and the structure of the graph itself, which shows independence. However, the lengths between pairs of nodes may not match the graph theoretic distances calculated in step 3.1.3. Here, we adopted Kamada and Kawai's refined spring model [120], minimizing the difference between energy states corresponds to the geometric and graph distances. The graph distances in this model are computed by and All-Pairs-Shortest-Path computation (Floyd-Warshall algorithm or Johnson's algorithm).

Clinical researchers usually need a quite large map. So, we use SFDP, a multi-level force-directed placement algorithm offering good layouts for large

graphs within reasonable running time. This algorithm uses a k-centers approximation to coarsen the original graph, followed by Kamada and Kawai's layout model. For implementation details in this algorithm, please refer [114]. We can compare the visual difference of FDP and SFDP on a top 800 diseases graph from Figure A.2 to Figure A.5. In Figure A.4, SFDP provides much clearer groups of nodes (like cities of countries in geographic maps) than that of FDP (Figure A.2). Obviously, SFDP algorithm makes it easy for doctors or physicians to understand the co-occurrence of prevalent diseases. In Figure A.3, the connection edges between pairs of diseases are so intricate that the layout is not informative for representation or research. In contrast, in Figure A.5, the diseases having high co-occurrence with Mixed Hyperlipidemia surround it without any edge intersection, which shows the relationship in a great way. Therefore, we recommend using SFDP for large map representation.

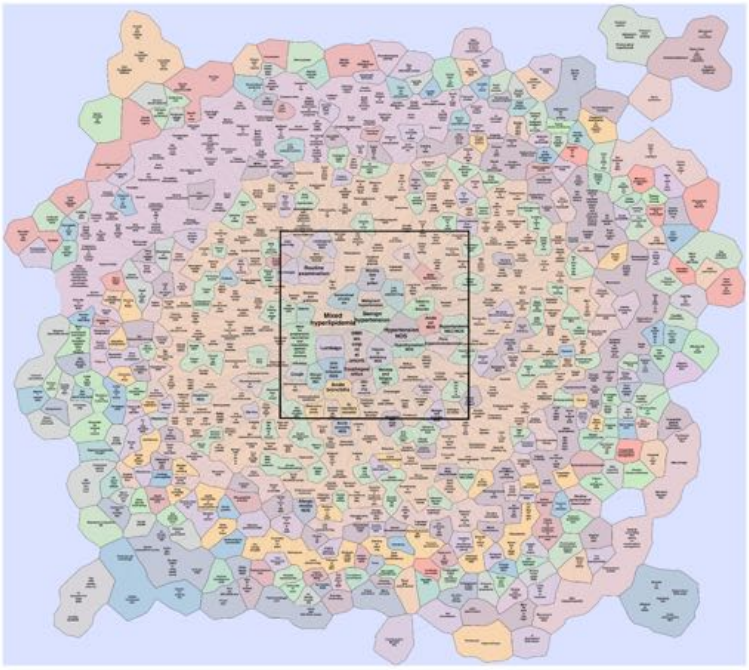


Figure A.2: A map of top 800 human diseases. The base map specification: FDP drawing algorithm, load factor: 1, edges enabled, coloring enabled.

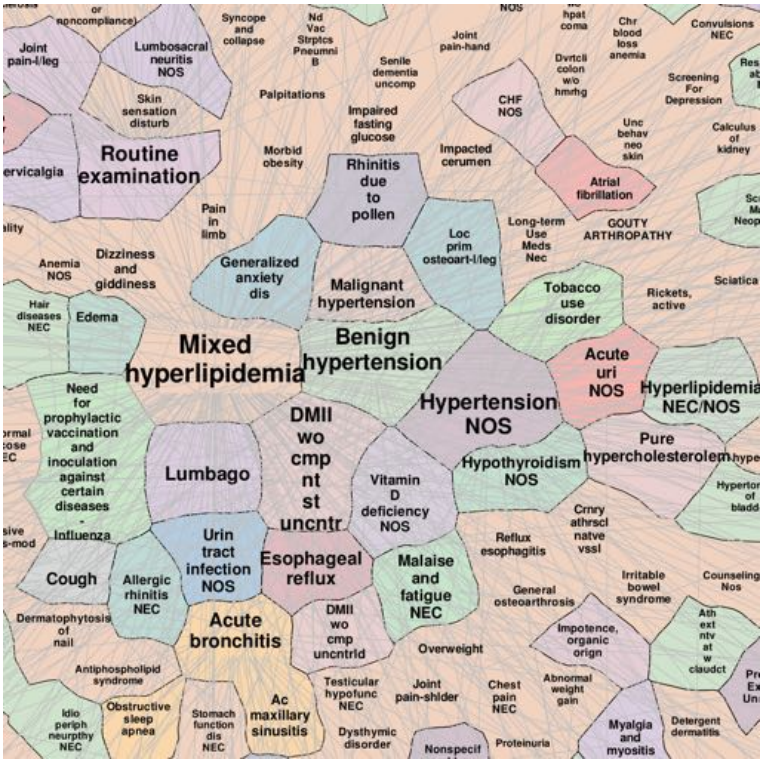


Figure A.3: The enlarged area within the black square in A.2.

3.2.2 Clustering:

Human diseases with high prevalence usually have dense connections among groups. We use a modularity based clustering method [107] to group nodes. It maximizes the modularity of the adjacent matrix of the graph via a spectral technique. The adjacent matrix modularity is defined as real edge numbers within groups minus the expected number of edges placed at random. Clusters can be detected based on computing the leading eigenvector of the adjacent matrix modularity and classifying nodes according to the signs in the eigenvector.

3.2.3 Geographic Mapping and Coloring:

To make a visually pleasing geographical map, we adopt a modified Voronoi diagram of the nodes as the mapping algorithm, which is described in [119].

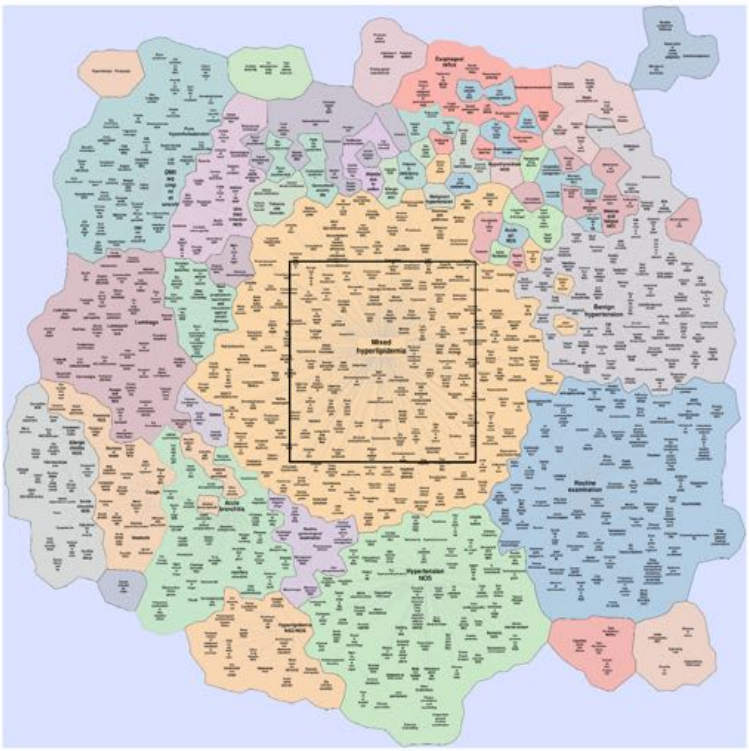


Figure A.4: A map of top 800 human diseases. The base map specification: SFDP drawing algorithm, load factor: 1, edges enabled, coloring enabled.

In this diagram, the boundaries of the clusters are more rounded than traditional Voronoi diagram, while some of the inner boundaries are kept artificially straight. In the last step, we color the countries (clusters) with ColorBrewer Theorem [121], which distinguishes adjacent clusters with two most different colors. For more information about the color distance function used in this method, please see details in [119].

A.3.3 Heat Map Generation

The heat map intensity in the overlay represents the significance of the information over certain nodes. There are three clinical information items: hospital, gender and age. The diameter of each heat spot is calculated by taking logarithm scale of normalized patient number density respective to

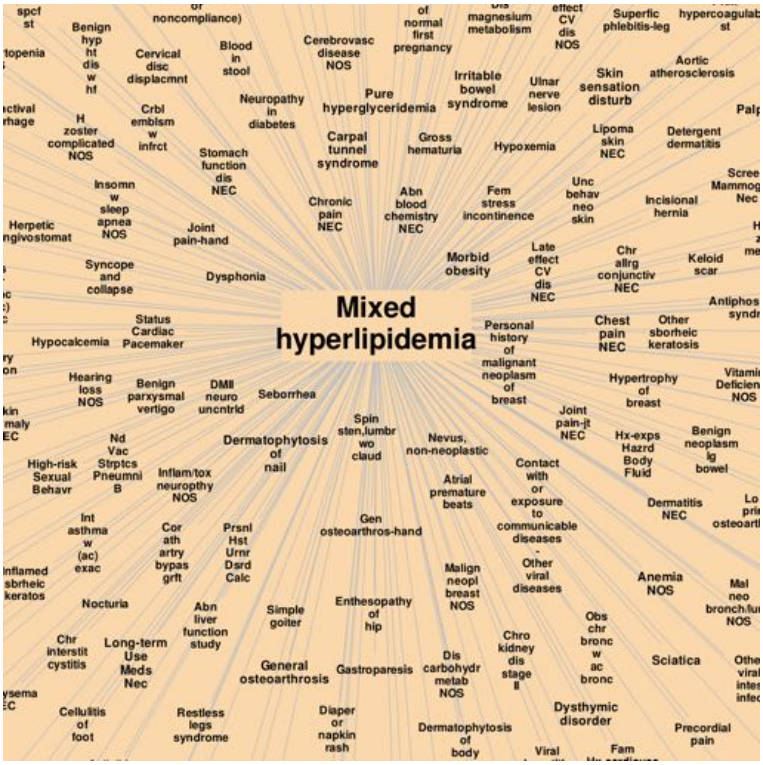


Figure A.5: The enlarged area within the black square in A.4.

the option within selected information type. Particularly, every node might have multiple highlight spots of information options, so we shift the heat spot centers by a little proportion in different directions to solve overlapping issue (as Figure A.1 shown). Solid semi-transparent circles are laid over the nodes as raster overlay.

A.3.4 Implementation

The system is implemented using Django Web application framework with modular design in Ubuntu 12.04. The four main modules in our system are: comorbidity graph generation, base map generation, heat map generation and map rendering. We render our SVG formatted base maps in web browser using AT&T's GraphViz system [106]. For the heat maps, we utilize heat map modules, together with zooming and panning tools from the open

source OpenLayers JavaScript library. The programming language used in the framework is Python 2.7.3.

A.4 Case Studies

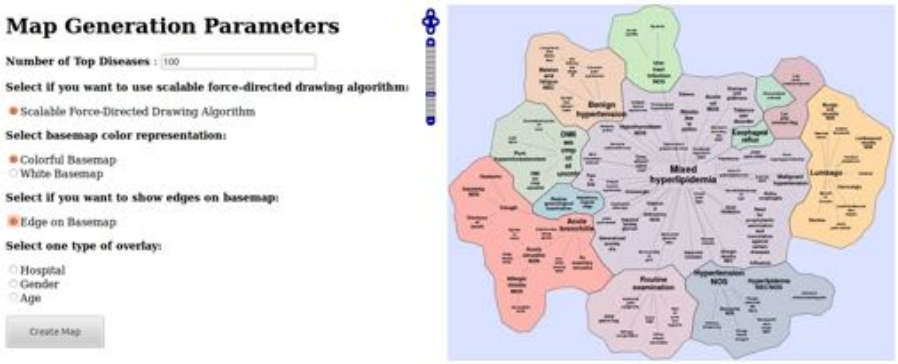


Figure A.6: A map of top 100 human diseases. The base map specification: SFDP drawing algorithm, load factor: 1, edges enabled, coloring enabled.

Two case studies focused on the base map and the heat map were done respectively to demonstrate the layout of disease co-occurrence relationships and clinical information highlight. There is a 'Map Generation Parameters' input user interface in our system, which is rendered in the web browser before map representation, as shown in the left part of Figure A.6.

In the first case study, we intend to get a colorful map of 100 prevalent diseases. In order to illustrate the most important relationship between diseases occurred frequently, we set the load factor for each node to be 1, which means only one edge emitting from each node in the graph. For better clustering, we choose SFDP for graph drawing algorithm. Since no heat map is needed in this case, we just leave the 'overlay' selection part blank, as shown in the user interface (Figure A.6). The result shows a hierarchical structure of a hundred prevalent diseases (extracted from the data set) that are grouped into 11 clusters. We can tell from the map that 'Mixed Hyperlipidemia' is the Top 1 human disease and it is the center of 9 sub-prevalent diseases. Obviously, the hierarchical structure improves our understanding of

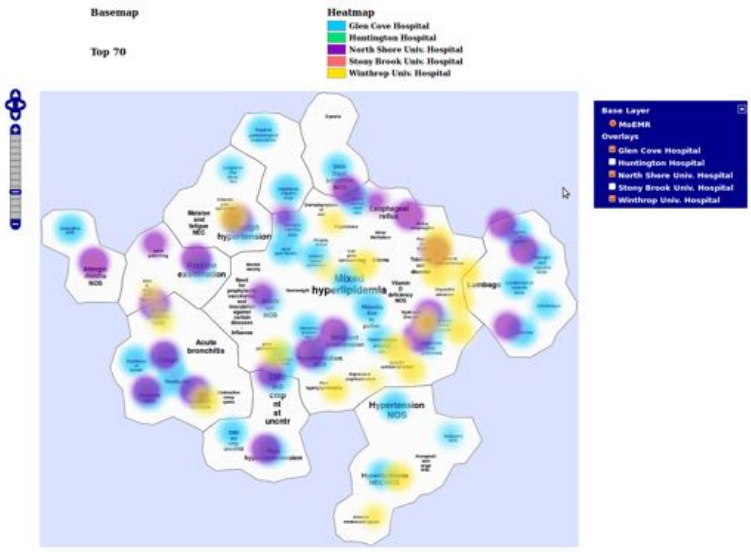


Figure A.7: Maps of top 70 human diseases. The base map specification: SFDP drawing algorithm, load factor: 1, edges disabled, coloring disabled. The heat map specification: the patient number density distributions of 5 large hospitals on Long Island, with Glen Cove Hospital, North Shore University Hospital and Winthrop University Hospital selected.

these one hundred diseases. It does make sense that mixed hyperlipidemia, hypertension and diabetes mellitus are highly related. For another example, in the red region (in the left bottom corner of the map), both acute bronchitis and acute sinusitis NOS are diseases of respiratory system.

In the second case study, the heat maps are clinical information overlays of hospital, age and gender respectively. In Figure A.7, we demonstrate three hospitals‘ heat maps (five hospitals altogether) over Top 70 Human Disease base map. The distribution of patient number density of every hospital is different, which can illustrate, to some degree, the clinical strength of hospitals or patient concentration near to hospitals. In Figure A.8, heat maps of three elder age ranges are overlaid on a base map of Top 60 Human Disease. In the map, the patient number density distribution is normalized in each age range. For example, in Figure A.8, people of elder than 72 years old cover more diseases than that of other age rangers. It is very interesting to do research on prevalent diseases in different age rangers. In Figure A.9, a

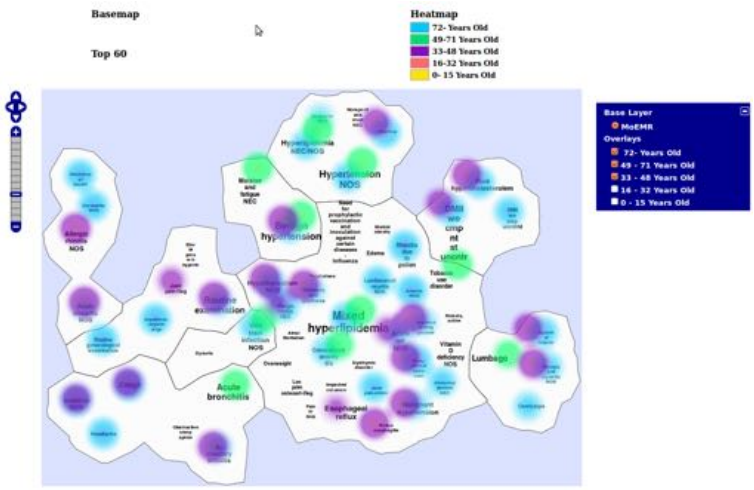


Figure A.8: Maps of top 60 human diseases. The base map specification: SFDP drawing algorithm, load factor: 1, edges disabled, coloring disabled. The heat map specification: the patient number density distributions of 5 age ranges, with >72 Years Old, 49-71 Years Old and 33-48 Years Old selected

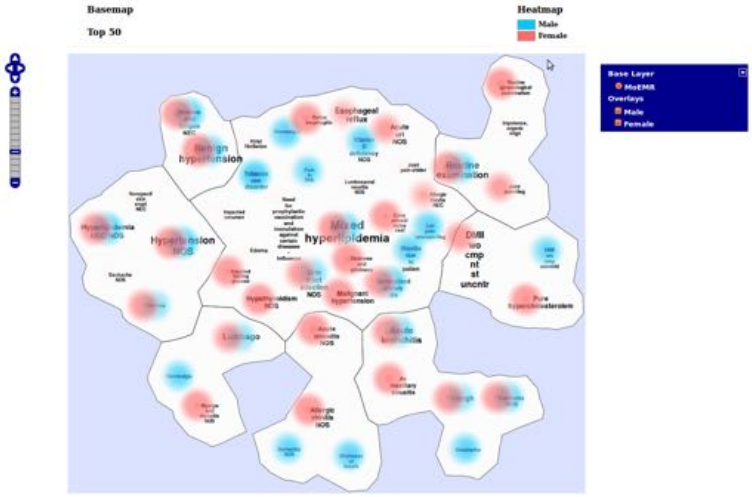


Figure A.9: Maps of top 50 human diseases. The base map specification: SFDP drawing algorithm, load factor: 1, edges disabled, coloring disabled. The heat map specification: the patient number density distributions of genders.

comparison of patient number density distribution from two genders is overlaid on a base map of Top 50 Human Disease. Despite Routine Gynecological Examination is exclusively for women, the information in this map driven by gender data shows significant trend of prevalent diseases among each gender. This visualization approach provides an efficient way for researchers or doctors to filter or determine gender-exclusive diseases.

A.5 Conclusion and Future Works

Our work proposes a practical framework for visual exploration of comorbidity between diseases. Experimental results show the goodness of SFDP graph drawing algorithm on large-scale data set, which also provide clinical meaning and nice visualization for researchers. Clinical information (hospital, age, gender) profiles are visualized by heat map overlay. In this paper, we have shown clinically valuable case studies that evaluate our system.

Future works could be considered in many promising directions to help medical personnel to arrive at more accurate diagnoses and treatments. For example, we can incorporate database technique with our visualization to standardize data entry queries and clinical data management. This work might need to be cooperated with physicians who are professional at EMR database. In addition, the flexibility of our system should be improved in respect of more types of clinical information overlay, comparable data visualization among options in a heat map and further statistical analysis. For comparable data visualization, we can solve this issue by normalizing patient number density in different groups. While this processing may lead to poor visual effect due to large distribution bias among different groups, we can create an option button in the user interface to control this functionality.

Appendix B

Digital Pathology Image Viewing Module

For visualization and algorithm implementation, 3D Slicer is an open-source medical image computing environment for clinical researchers and it is well designed to extend modules for programmers. I developed modules in 3D Slicer to read/write, visualize and process high-resolution images used in digital pathology, which can occupy tens of gigabytes and cannot be comfortably uncompressed into RAM. Slicer supports three types of modules: command line interface (CLI), scripted modules and loadable modules. I used scripted modules as they are recommended for fast prototyping and custom workflow development. Like my standalone application, a scripted module in 3D Slicer consists on widget part and logic part. In Slicer, the global repository for all data model is MRML scene. Each MRML scene contains list of nodes (display and storage) and relationship between nodes (references and hierarchies). The widget class defines the module user interface and launches processing methods implemented in the logic class. It is important that the widget class also keeps user interface and nodes in synchronization by observing MRML nodes to get change notifications. The logic class is the reusable core of the processing program. The logic module should be separated from the widget class and MRML nodes in object oriented design. I will learn and comply rules and conventions in 3D Slicer during the development to accomplish an efficient scripted module.

A standalone application has been written to import and visualize the large two-dimensional histopathological images via interactive input of multi-

ple parameters. For visualization implementation, I used PyQt and openslide-python library. The interactive front panel is shown in the following Figure B.1. The image region being visualized is defined by multiple parameters by users. As shown in the figure, after loading specific image file with .svs suffix, the filepath, filename and image meta information will be displayed. According to the dimension, level number and down sample rate on each level, users can input their desired level, start point position, width and height of the region. Then the desired image region can be shown on the left scene by simply clicking the 'Show Image' button. The image input mechanism is implemented by utilizing the methods in openslide and openslie-python libraries. The image data were converted into numpy array and then were displayed in the interface using methods in QGraphicsScene.

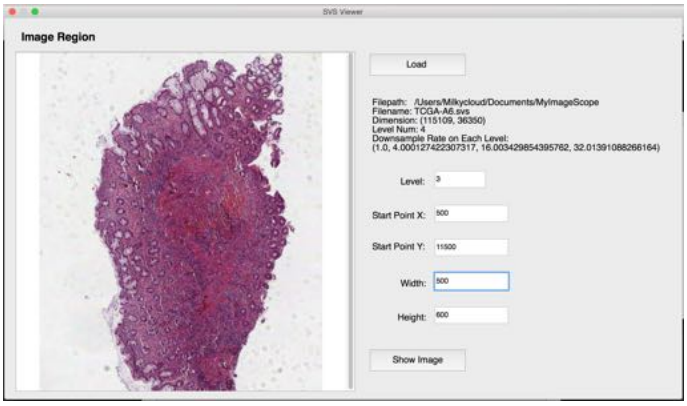


Figure B.1: The interactive panel of my histopathology image visualization tool. The source code is stored in my Github account repository: https://github.com/naiyunzhou/GSOC2016.

Future possible biomedical engineering or informatics research could be considered in many promising directions to help medical personnel to arrive at more accurate diagnoses and treatments. For example, we can incorporate database technique with our visualization to standardize data entry queries and clinical data management. This work might need to be co-operated with physicians who are professional at EMR database. In addition, the flexibility of our system should be improved in respect of more types of clinical information overlay, comparable data visualization among options in a heat map and further statistical analysis.

Bibliography

[1] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal, "Cancer statistics, 2015," *CA: a cancer journal for clinicians*, vol. 65, no. 1, pp. 5–29, 2015.

[2] Peter Boyle, Bernard Levin, et al., *World cancer report 2008.*, IARC Press, International Agency for Research on Cancer, 2008.

[3] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener, "Histopathological image analysis: a review," *Biomedical Engineering, IEEE Reviews in*, vol. 2, pp. 147–171, 2009.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[5] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1717–1724.

[6] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," *arXiv preprint arXiv:1409.4842*, 2014.

[7] H Fox, "Is h&e morphology coming to an end?," *Journal of clinical pathology*, vol. 53, no. 1, pp. 38–40, 2000.

[8] Ruixia Zhou, Elizabeth H Hammond, and Dennis L Parker, "A multiple wavelength algorithm in color image analysis and its applications in

stain decomposition in microscopy images," *Medical physics*, vol. 23, no. 12, pp. 1977–1986, 1996.

[9] Bernd Jahne and Bernd Jaehne, *Practical handbook on image processing for scientific applications*, CRC Press, Inc., 1995.

[10] Arnout C Ruifrok and Dennis A Johnston, "Quantification of histochemical staining by color deconvolution.," *Analytical and quantitative cytology and histology/the International Academy of Cytology [and] American Society of Cytology*, vol. 23, no. 4, pp. 291–299, 2001.

[11] Andrew Rabinovich, Sameer Agarwal, Casey Laris, Jeffrey H Price, and Serge J Belongie, "Unsupervised color decomposition of histologically stained tissue samples," in *Advances in neural information processing systems*, 2003, p. None.

[12] Milan Gavrilovic, Jimmy C Azar, Joakim Lindblad, Carolina Wahlby, Ewert Bengtsson, Christoph Busch, and Ingrid B Carlbom, "Blind color decomposition of histological images," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 6, pp. 983–994, 2013.

[13] LG Fulford, DF Easton, JS Reis-Filho, A Sofronis, CE Gillett, SR Lakhani, and A Hanby, "Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast," *Histopathology*, vol. 49, no. 1, pp. 22–34, 2006.

[14] Kien Nguyen, Bikash Sabata, and Anil K Jain, "Prostate cancer grading: Gland segmentation and structural features," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 951–961, 2012.

[15] Shivang Naik, Scott Doyle, Shannon Agner, Anant Madabhushi, Michael Feldman, and John Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE, 2008, pp. 284–287.

[16] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[17] Sahirzeeshan Ali and Anant Madabhushi, "An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery," *Medical Imaging, IEEE Transactions On*, vol. 31, no. 7, pp. 1448–1460, 2012.

[18] Scott Doyle, Mark Hwang, Kinsuk Shah, Anant Madabhushi, Michael Feldman, and John Tomas, "Automated grading of prostate cancer using architectural and textural image features," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*. IEEE, 2007, pp. 1284–1287.

[19] Ajay Basavanhally, Shridar Ganesan, Natalie Shih, Carolyn Mies, Michael Feldman, John Tomaszewski, and Anant Madabhushi, "A boosted classifier for integrating multiple fields of view: Breast cancer grading in histopathology," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 125–128.

[20] Grigory Begelman, Michael Pechuk, Ehud Rivlin, and Edmond Sabo, "System for computer-aided multiresolution microscopic pathology diagnostics," in *Computer Vision Systems, 2006 ICVS'06. IEEE International Conference on*. IEEE, 2006, pp. 16–16.

[21] Xin Liu, Deanna L Langer, Masoom A Haider, Yongyi Yang, Miles N Wernick, and İmam Şamil Yetik, "Prostate cancer segmentation with simultaneous estimation of markov random field parameters and class," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 6, pp. 906–915, 2009.

[22] J Monaco, J Tomaszewski, M Feldman, Mehdi Moradi, Parvin Mousavi, Alexander Boag, Chris Davidson, Purang Abolmaesumi, and Anant Madabhushi, "Detection of prostate cancer from whole-mount histology images using markov random fields," in *Workshop on Microscopic Image Analysis with Applications in Biology (in conjunction with MICCAI)*. Citeseer, 2008.

[23] Jun Xu, James P Monaco, and Anant Madabhushi, "Markov random field driven region-based active contour model (maracel): application to medical image segmentation," in *Medical Image Computing and*

Computer-Assisted Intervention–MICCAI 2010, pp. 197–204. Springer, 2010.

[24] Yoshua Bengio, Ian J Goodfellow, and Aaron Courville, "Deep learning," *An MIT Press book in preparation. Draft chapters available at http://www. iro. umontreal. ca/ bengioy/dlbook*, 2015.

[25] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[26] James Monaco, J Hipp, D Lucas, S Smith, U Balis, and Anant Madabhushi, "Image segmentation with implicit color standardization using spatially constrained expectation maximization: Detection of nuclei," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 365–372. Springer, 2012.

[27] Christopher W Elston, Ian O Ellis, et al., "Pathological prognostic factors in breast cancer. i. the value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.

[28] Farzad Ghaznavi, Andrew Evans, Anant Madabhushi, and Michael Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," *Annual Review of Pathology: Mechanisms of Disease*, vol. 8, pp. 331–359, 2013.

[29] Lei He, L Rodney Long, Sameer Antani, and George R Thoma, "Histology image analysis for carcinoma detection and grading," *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 538–556, 2012.

[30] Siamak Daneshmand, Marcus L Quek, John P Stein, Gary Lieskovsky, Jie Cai, Jacek Pinski, Eila C Skinner, and Donald G Skinner, "Prognosis of patients with lymph node positive prostate cancer following radical prostatectomy: long-term results," *The Journal of urology*, vol. 172, no. 6, pp. 2252–2255, 2004.

[31] Jonathan I Epstein, "An update of the gleason grading system," *The Journal of urology*, vol. 183, no. 2, pp. 433–440, 2010.

[32] Jonathan L Wright, Claudia A Salinas, Daniel W Lin, Suzanne Kolb, Joseph Koopmeiners, Ziding Feng, and Janet L Stanford, "Prostate cancer specific mortality and gleason 7 disease differences in prostate cancer outcomes between cases with gleason 4+ 3 and gleason 3+ 4 tumors in a population based cohort," *The Journal of urology*, vol. 182, no. 6, pp. 2702–2707, 2009.

[33] Shoshana Ginsburg, George Lee, Sahirzeeshan Ali, and Anant Madabhushi, "Feature importance in nonlinear embeddings (fine): Applications in digital pathology," 2010.

[34] Amrita Mohanty, Zameer Muzaffar Mir, S Rajkumar, and Puja Bardhan, "Analysis of color images using cluster based segmentation techniques," *Analysis*, vol. 79, no. 2, 2013.

[35] Godwin Avwioro, "Histochemical uses of haematoxylin—a review," *JPCS*, vol. 1, pp. 24–34, 2011.

[36] David F Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of computation*, vol. 24, no. 111, pp. 647–656, 1970.

[37] Rachel Sparks and Anant Madabhushi, "Explicit shape descriptors: Novel morphologic features for histopathology classification," *Medical image analysis*, vol. 17, no. 8, pp. 997–1009, 2013.

[38] Lena Gorelick, Olga Veksler, Mena Gaed, Jairo Alejandro Gomez, Madeleine Moussa, Glenn Bauman, Aaron Fenster, and Aaron D Ward, "Prostate histopathology: learning tissue component histograms for cancer detection and classification," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 10, pp. 1804–1818, 2013.

[39] George Lee, Rachel Sparks, Sahirzeeshan Ali, Natalie NC Shih, Michael D Feldman, Elaine Spangler, Timothy Rebbeck, John E Tomaszewski, and Anant Madabhushi, "Co-occurring gland angularity in localized subgraphs: Predicting biochemical recurrence in intermediate-risk prostate cancer patients," 2014.

[40] Haibo Wang, Angel Cruz-Roa, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, and

Anant Madabhushi, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *Journal of Medical Imaging*, vol. 1, no. 3, pp. 034003–034003, 2014.

[41] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pp. 411–418. Springer, 2013.

[42] Angel Alfonso Cruz-Roa, John Edison Arevalo Ovalle, Anant Madabhushi, and Fabio Augusto González Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, pp. 403–410. Springer, 2013.

[43] Yan Xu, Zhipeng Jia, Yuqing Ai, Fang Zhang, Maode Lai, Eric I Chang, et al., "Deep convolutional activation features for large scale brain tumor histopathology image classification and segmentation," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 947–951.

[44] Angel Cruz-Roa, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2014, pp. 904103–904103.

[45] James MacQueen et al., "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA., 1967, vol. 1, pp. 281–297.

[46] Artem Babenko, Anton Slesarev, Alexandr Chigorin, and Victor Lempitsky, "Neural codes for image retrieval," in *Computer Vision–ECCV 2014*, pp. 584–599. Springer, 2014.

[47] `http://miccai.cloudapp.net/competitions/57`.

[48] Jason Tseung, "Robbins and cotran pathologic basis of disease," 2005.

[49] Andrew H Fischer, Kenneth A Jacobson, Jack Rose, and Rolf Zeller, "Hematoxylin and eosin staining of tissue and cell sections," *Cold Spring Harbor Protocols*, vol. 2008, no. 5, pp. pdb–prot4986, 2008.

[50] `http://www.pathpedia.com/education/eatlas/histopathology/breast.aspx`.

[51] O.S. Eldin, *The Skilful Pathologist: Pathology Concepts in Histopathology Diagnosis*, Lulu.com, 2011.

[52] Pranab Dey, "Cancer nucleus: morphology and beyond," *Diagnostic cytopathology*, vol. 38, no. 5, pp. 382–390, 2010.

[53] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.

[54] Youyi Song, Ling Zhang, Siping Chen, Dong Ni, Baiying Lei, and Tianfu Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2421–2433, 2015.

[55] Fuyong Xing, Yuanpu Xie, and Lin Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 550–566, 2016.

[56] Andrew Janowczyk and Anant Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *Journal of pathology informatics*, vol. 7, 2016.

[57] Xipeng Pan, Lingqiao Li, Huihua Yang, Zhenbing Liu, Jinxin Yang, Lingling Zhao, and Yongxian Fan, "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88–99, 2017.

[58] Hao Chen, Xiaojuan Qi, Lequan Yu, Qi Dou, Jing Qin, and Pheng-Ann Heng, "Dcan: Deep contour-aware networks for object instance

segmentation from histology images," *Medical image analysis*, vol. 36, pp. 135–146, 2017.

[59] Le Hou, Vu Nguyen, Dimitris Samaras, Tahsin M Kurc, Yi Gao, Tianhao Zhao, and Joel H Saltz, "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," *arXiv preprint arXiv:1704.00406*, 2017.

[60] `https://en.wikipedia.org/wiki/Student%27s_t-test`.

[61] Shivang Naik, Scott Doyle, Michael Feldman, John Tomaszewski, and Anant Madabhushi, "Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information," in *MIAAB workshop*. Citeseer, 2007, pp. 1–8.

[62] Petros S Karvelis, Dimitrios I Fotiadis, Ioannis Georgiou, and Marika Syrrou, "A watershed based segmentation method for multispectral chromosome images classification," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 3009–3012.

[63] Sokol Petushi, Fernando U Garcia, Marian M Haber, Constantine Katsinis, and Aydin Tozeren, "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC medical imaging*, vol. 6, no. 1, pp. 1, 2006.

[64] Donald L Weaver, David N Krag, Edward A Manna, Taka Ashikaga, Seth P Harlow, and Kenneth D Bauer, "Comparison of pathologist-detected and automated computer-assisted image analysis detected sentinel lymph node micrometastases in breast cancer," *Modern pathology*, vol. 16, no. 11, pp. 1159–1163, 2003.

[65] Lin Yang, Peter Meer, and David J Foran, "Unsupervised segmentation based on robust estimation and color active contour models," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 3, pp. 475–486, 2005.

[66] Carolina Wählby, I-M SINTORN, Fredrik Erlandsson, Gunilla Borgefors, and Ewert Bengtsson, "Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections," *Journal of Microscopy*, vol. 215, no. 1, pp. 67–76, 2004.

[67] Metin N Gurcan, Tony Pan, Hiro Shimada, and Joel Saltz, "Image analysis for neuroblastoma classification: segmentation of cell nuclei," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 4844–4847.

[68] Lee AD Cooper, Jun Kong, David A Gutman, Fusheng Wang, Jingjing Gao, Christina Appin, Sharath Cholleti, Tony Pan, Ashish Sharma, Lisa Scarpace, et al., "Integrated morphologic analysis for the identification and characterization of disease subtypes," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 317–323, 2012.

[69] Jun Kong, Lee AD Cooper, Fusheng Wang, Jingjing Gao, George Teodoro, Lisa Scarpace, Tom Mikkelsen, Matthew J Schniederjan, Carlos S Moreno, Joel H Saltz, et al., "Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates," *PloS one*, vol. 8, no. 11, pp. e81049, 2013.

[70] Mitko Veta, Paul J van Diest, Robert Kornegoor, André Huisman, Max A Viergever, and Josien PW Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PloS one*, vol. 8, no. 7, pp. e70221, 2013.

[71] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2957, oct 2010.

[72] Luc Devroye, "Sample-based non-uniform random variate generation," in *Proceedings of the 18th conference on Winter simulation*. ACM, 1986, pp. 260–265.

[73] Eldad Haber, Tauseef Rehman, and Allen Tannenbaum, "An efficient numerical method for the solution of the l_2 optimal mass transfer problem," *SIAM Journal on Scientific Computing*, vol. 32, no. 1, pp. 197–211, 2010.

[74] Yi Gao, Liangjia Zhu, Sylvain Bouix, and Allen Tannenbaum, "Interpolation of longitudinal shape and image data via optimal mass transport," in *Proceedings of SPIE*, 2014, p. 90342X.

[75] Rainer E Burkard, Mauro Dell'Amico, and Silvano Martello, *Assignment Problems, Revised Reprint*, Siam, 2009.

[76] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[77] Sarang C. Joshi, Michael I. Miller, and Ulf Grenander, "On the geometry and shape of brain sub-manifolds," *IJPRAI*, vol. 11, no. 8, pp. 1317–1343, 1997.

[78] Samuel Dambreville, Yogesh Rathi, and Allen Tannenbaum, "A framework for image segmentation using shape models and kernel space shape priors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 8, pp. 1385–1399, 2008.

[79] Sam T Roweis and Lawrence K Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[80] T.T. Cormen, C.E. Leiserson, and R.L. Rivest, *Introduction to algorithms*, MIT Press, 2001.

[81] David G Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces," *Bulletin of the London Mathematical Society*, vol. 16, no. 2, pp. 81–121, 1984.

[82] Kaleem Siddiqi and Stephen Pizer, *Medial representations: mathematics, algorithms and applications*, vol. 37, Springer Science & Business Media, 2008.

[83] Andrew Blake, Pushmeet Kohli, and Carsten Rother, *Markov random fields for vision and image processing*, Mit Press, 2011.

[84] Alexei A Efros and Thomas K Leung, "Texture synthesis by non-parametric sampling," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. IEEE, 1999, vol. 2, pp. 1033–1038.

[85] Shawn Lankton and Allen Tannenbaum, "Localizing region-based active contours," *IEEE transactions on image processing*, vol. 17, no. 11, pp. 2029–2039, 2008.

[86] Yizong Cheng, "Mean shift, mode seeking, and clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[87] Dorin Comaniciu and Peter Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[88] Luis Ibanez, William Schroeder, Lydia Ng, and Josh Cates, "The itk software guide," 2005.

[89] Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, Mahadev Satyanarayanan, et al., "Openslide: A vendor-neutral software foundation for digital pathology," *Journal of pathology informatics*, vol. 4, no. 1, pp. 27, 2013.

[90] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, 2014.

[91] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2015.

[92] Li-Yi Wei and Marc Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 479–488.

[93] Javier Portilla and Eero P Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International journal of computer vision*, vol. 40, no. 1, pp. 49–70, 2000.

[94] Alexei A Efros and William T Freeman, "Image quilting for texture synthesis and transfer," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001, pp. 341–346.

[95] Vivek Kwatra, Arno Schödl, Irfan Essa, Greg Turk, and Aaron Bobick, "Graphcut textures: image and video synthesis using graph cuts," in *ACM Transactions on Graphics (ToG)*. ACM, 2003, vol. 22, pp. 277–286.

[96] Li-Yi Wei, Sylvain Lefebvre, Vivek Kwatra, and Greg Turk, "State of the art in example-based texture synthesis," in *Eurographics 2009, State of the Art Report, EG-STAR*. Eurographics Association, 2009, pp. 93–117.

[97] Leon Gatys, Alexander S Ecker, and Matthias Bethge, "Texture synthesis using convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 262–270.

[98] John D'Errico, "An open source matlab image inpainting package," 2012.

[99] Nasir Rajpoot, "The challenge on gland segmentation in histology images," 2015.

[100] Luis Ibanez, William Schroeder, Lydia Ng, and Josh Cates, "The itk software guide," 2003.

[101] Francesco Folino, Clara Pizzuti, and Maria Ventura, "A comorbidity network approach to predict disease risk," in *Proceedings of the First International Conference on Information Technology in Bio- and Medical Informatics*, Berlin, Heidelberg, 2010, ITBAM'10, pp. 102–109, Springer-Verlag.

[102] Chris van Weel and François G Schellevis, "Comorbidity and guidelines: conflicting interests," *The Lancet*, vol. 367, no. 9510, pp. 550–551, 2006.

[103] Enrico Capobianco and Pietro Liò, "Comorbidity networks: beyond disease correlations," *Journal of Complex Networks*, 2015.

[104] Mohammad Moni and Pietro Lio, "comor: a software for disease comorbidity risk assessment," *Journal of Clinical Bioinformatics*, vol. 4, no. 1, pp. 8, 2014.

[105] Daniel Fried and Stephen G Kobourov, "Maps of computer science," in *Pacific Visualization Symposium (PacificVis), 2014 IEEE*. IEEE, 2014, pp. 113–120.

[106] John Ellson, Emden Gansner, Lefteris Koutsofios, Stephen C North, and Gordon Woodhull, "Graphviz—open source graph drawing tools," in *Graph Drawing*. Springer, 2002, pp. 483–484.

[107] Andreas Noack, "Modularity clustering is force-directed layout," *Physical Review E*, vol. 79, no. 2, pp. 026102, 2009.

[108] Krist Wongsuphasawat and David Gotz, "Outflow: Visualizing patient flow by symptoms and outcome," in *IEEE VisWeek Workshop on Visual Analytics in Healthcare, Providence, Rhode Island, USA*, 2011.

[109] Zhiyuan Zhang, Arunesh Mittal, Supriya Garg, Alexander E Dimitriyadi, IV Ramakrishnan, Rong Zhao, Asa Viccellio, and Klaus Mueller, "A visual analytics framework for emergency room clinical encounters," in *IEEE Workshop on Visual Analytics in Health Care*, 2010.

[110] Peng Gang Sun, Lin Gao, and Shan Han, "Prediction of human disease-related gene clusters by clustering analysis," *International journal of biological sciences*, vol. 7, no. 1, pp. 61, 2011.

[111] Kwang-Il Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási, "The human disease network," *Proceedings of the National Academy of Sciences*, vol. 104, no. 21, pp. 8685–8690, 2007.

[112] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo, "Network medicine: a network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, no. 1, pp. 56–68, 2011.

[113] Thomas MJ Fruchterman and Edward M Reingold, "Graph drawing by force-directed placement," *Software: Practice and experience*, vol. 21, no. 11, pp. 1129–1164, 1991.

[114] David Harel and Yehuda Koren, "A fast multi-scale method for drawing large graphs," in *Graph drawing*. Springer, 2001, pp. 183–196.

[115] Anna Chmiel, Peter Klimek, and Stefan Thurner, "Spreading of diseases through comorbidity networks across life and gender," *New Journal of Physics*, vol. 16, no. 11, pp. 115013, 2014.

[116] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu, "Using of jaccard coefficient for keywords similarity," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2013, vol. 1, p. 6.

[117] Yifan Hu, "Efficient, high-quality force-directed graph drawing," *Mathematica Journal*, vol. 10, no. 1, pp. 37–71, 2005.

[118] Mark EJ Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[119] Emden R Gansner, Yifan Hu, and Stephen Kobourov, "Gmap: Visualizing graphs and clusters as maps," in *Pacific Visualization Symposium (PacificVis), 2010 IEEE*. IEEE, 2010, pp. 201–208.

[120] Tomihisa Kamada and Satoru Kawai, "An algorithm for drawing general undirected graphs," *Information processing letters*, vol. 31, no. 1, pp. 7–15, 1989.

[121] Mark Harrower and Cynthia A Brewer, "Colorbrewer. org: an online tool for selecting colour schemes for maps," *The Cartographic Journal*, vol. 40, no. 1, pp. 27–37, 2003.