DEEP LEARNING FOR DIGITAL PATHOLOGY

A DISSERTATION SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

By
Can Taylan Sarı
November 2020

DEEP LEARNING FOR DIGITAL PATHOLOGY
By Can Taylan Sarı
November 2020

We certify that we have read this dissertation and that in our opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Çiğdem Gündüz Demir(Advisor)

Pınar Duygulu Şahin

Hamdi Dibeklioğlu

Alptekin Temizel

Abdullah Ercüment Çiçek

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

DEEP LEARNING FOR DIGITAL PATHOLOGY

Can Taylan Sarı
Ph.D. in Computer Engineering
Advisor: Çiğdem Gündüz Demir
November 2020

Histopathological examination is today's gold standard for cancer diagnosis and grading. However, this task is time consuming and prone to errors as it requires detailed visual inspection and interpretation of a histopathological sample provided on a glass slide under a microscope by an expert pathologist. Low-cost and high-technology whole slide digital scanners produced in recent years have eliminated the disadvantages of physical glass slide samples by digitizing histopathological samples and relocating them to digital media. Digital pathology aims at alleviating the problems of traditional examination approaches by providing auxiliary computerized tools that quantitatively analyze digitized histopathological images.

Traditional machine learning methods have proposed to extract handcrafted features from histopathological images and to use these features in the design of a classification or a segmentation algorithm. The performance of these methods mainly relies on the features that they use, and thus, their success strictly depends on the ability of these features to successfully quantify the histopathology domain. More recent studies have employed deep architectures to learn expressive and robust features directly from images avoiding complex feature extraction procedures of traditional approaches. Although deep learning methods perform well in many classification and segmentation problems, convolutional neural networks that they frequently make use of require annotated data for training and this makes it difficult to utilize unannotated data that cover the majority of the available data in the histopathology domain.

This thesis addresses the challenges of traditional and deep learning approaches by incorporating unsupervised learning into classification and segmentation algorithms for feature extraction and training regularization purposes in the histopathology domain. As the first contribution of this thesis, the first study presents a new unsupervised feature extractor for effective representation and classification of histopathological tissue images. This study introduces a deep belief network to quantize the salient subregions, which are identified with

domain-specific prior knowledge, by extracting a set of features directly learned on image data in an unsupervised way and uses the distribution of these quantizations for image representation and classification. As its second contribution, the second study proposes a new regularization method to train a fully convolutional network for semantic tissue segmentation in histopathological images. This study relies on the benefit of unsupervised learning, in the form of image reconstruction, for network training. To this end, it puts forward an idea of defining a new embedding, which is generated by superimposing an input image on its segmentation map, that allows uniting the main supervised task of semantic segmentation and an auxiliary unsupervised task of image reconstruction into a single one and proposes to learn this united task by a generative adversarial network. We compare our classification and segmentation methods with traditional machine learning methods and the state-of-the-art deep learning algorithms on various histopathological image datasets. Visual and quantitative results of our experiments demonstrate that the proposed methods are capable of learning robust features from histopathological images and provides more accurate results than their counterparts.

Keywords: Deep learning, feature learning, training regularization, image embedding, generative adversarial networks, semantic segmentation, digital pathology, automated cancer diagnosis, histopathological image analysis.

ÖZET

DİJİTAL PATOLOJİ İÇİN DERİN ÖĞRENME

Can Taylan Sarı
Bilgisayar Mühendisliği, Doktora
Tez Danışmanı: Çiğdem Gündüz Demir
Kasım 2020

Histopatolojik değerlendirme, kanser teşhisi ve derecelendirmesi için günümüzde kullanılan araçtır. Öte yandan, bu değerlendirme, cam slayt üzerindeki histopatolojik numunenin uzman bir patolog tarafından mikroskop altında ayrıntılı olarak incelenmesini ve yorumlanmasını gerektirdiğinden, zaman alıcı ve hatalara açık bir işlemdir. Son yıllarda üretilen düşük maliyetli ve yüksek teknolojili tam slayt dijital tarayıcılar, histopatolojik örnekleri dijital ortama aktararak, fiziksel cam slayt örneklerin dezavantajlarını ortadan kaldırmaktadır. Dijital patoloji, dijitalleştirilmiş histopatolojik görüntüleri nicel olarak analiz eden yardımcı bilgisayarlı araçlar sağlayarak geleneksel inceleme yaklaşımlarının sorunlarını azaltmayı amaçlamaktadır.

Geleneksel makine öğrenmesi yöntemleri, histopatolojik görüntülerden manuel tanımlanmış öznitelikler çıkarmayı ve bu öznitelikleri bir sınıflandırma veya bölütleme algoritması tasarımında kullanmayı önermektedir. Bu yöntemlerin performansı esas olarak kullandıkları özniteliklere dayanmaktadır ve bu nedenle, bu yöntemlerin başarıları, kullandıkları özniteliklerin histopatoloji alanını başarılı bir şekilde temsil etme yeteneklerine bağlıdır. Son yıllarda önerilen çalışmalar, geleneksel yaklaşımların karmaşık öznitelik çıkarma prosedürlerinden kaçınarak, açıklayıcı ve gürbüz öznitelikleri doğrudan görüntülerden öğrenmek için derin mimariler kullanmaktadır. Derin öğrenme yöntemleri birçok sınıflandırma ve bölütleme probleminde iyi performans gösterse de, sıklıkla kullandıkları evrişimsel sinir ağları eğitim için etiketlenmiş verilere ihtiyaç duymaktadır ve bu da, histopatoloji alanındaki mevcut verilerin çoğunu kapsayan etiketlenmemiş verilerin kullanılmasını zor hale getirmektedir.

Bu tez, geleneksel yöntemlerin ve derin öğrenme yaklaşımlarının sorunlarını, denetimsiz öğrenmenin öznitelik çıkarma ve eğitim düzenleme amaçları için sınıflandırma ve bölütleme algoritmalarına dahil edilmesiyle ele alınmaktadır. Tezin birinci katkısı olarak sunulan ilk çalışma, histopatolojik doku görüntülerinin etkili bir şekilde temsil edilmesi ve sınıflandırılması için yeni bir denetimsiz

öznitelik çıkarıcı sunmaktadır. Bu çalışmada, alana özgü ön bilgilerle tanımlanan önemli alt bölgelerden öznitelikler çıkarmak amacıyla, denetimsiz bir derin inanç ağı eğitilmiş ve bu eğitim sonucunda elde edilen özniteliklerin dağılımı, görüntü gösterimi ve sınıflandırması için kullanılmıştır. Tezin ikinci katkısı olarak sunulan diğer çalışmada, histopatolojik doku görüntülerinde semantik doku bölütlemesi için, tam bağlantılı bir evrişimsel ağ eğitmek amacıyla yeni bir düzenleme yöntemi önerilmektedir. Bu çalışma, denetimsiz öğrenmeyi, önerilen ağ modelinin eğitimini düzenlemek için, girdi görüntülerinin yeniden yapılandırılması şeklinde kullanmaktadır. Bu amaçla, bölütleme haritası ile girdi görüntüsünün üst üste bindirilmesiyle oluşturulan yeni bir yerleştirme tanımlanmaktadır. Önerilen bu yerleştirme yöntemi sayesinde, semantik bölütlemeyi temsil eden ana denetimli görev ile görüntüyü yeniden yapılandırmanın temsil ettiği yardımcı denetimsiz görevin tek bir görevde birleştirilmesi ve oluşturulan bu birleşik görevin, bir üretken çekişmeli ağ ile öğrenilmesi amaçlanmaktadır. Önerilen sınıflandırma ve bölütleme yöntemleri, geleneksel makine öğrenmesi yöntemleri ve güncel derin öğrenme algoritmalarıyla, farklı histopatolojik görüntü veri kümeleri kullanılarak karşılaştırılmıştır. Deneyler sonucunda elde edilen görsel ve nicel sonuçlar, önerilen yöntemlerin histopatolojik görüntülerden gürbüz öznitelikler öğrenebildiğini ve karşılaştırılan yöntemlerden daha doğru sonuçlar ürettiğini ortaya koymaktadır.

Anahtar sözcükler: Derin öğrenme, öznitelik öğrenme, eğitim düzenleme, görüntü yerleştirme, üretken çekişmeli ağlar, anlamsal bölütleme, dijital patoloji, otomatik kanser teşhisi, histopatolojik görüntü analizi.

Acknowledgement

I would like to dedicate this thesis to my wife, Aylin, for her endless support and unrequited love. None of this would have been possible without her.

First and foremost, I would like to express my deepest and sincerest gratitude to my advisor, Assoc. Prof. Çiğdem Gündüz Demir for giving me the opportunity of studying under her supervision and providing invaluable academic guidance throughout my doctorate education. This thesis was completed thanks to her excellent field knowledge and the patience and encouragement she has shown me in my hardest times.

I would like to thank my tracking committee members; Prof. Dr. Pınar Duygulu Şahin and Asst. Prof. Dr. Hamdi Dibeklioğlu for their precious time and valuable feedbacks. I would also like to thank to Prof. Dr. Alptekin Temizel and Asst. Prof. Abdullah Ercüment Çiçek for being members of my Ph.D. thesis jury and reviewing on this thesis.

I would like to thank Prof. Dr. Cenk Sökmensüer for his guidance in pathology and providing datasets for the studies in this thesis.

I would like to thank my dear friends Fuat and Arif, whom I have always felt by my side with their academic and social support during my doctoral course period, qualifying exam and thesis studies.

Finally, I sincerely thank my deceased father, mother and brothers for their faith and support throughout my academic life since primary school.

Part of this work is reprinted, with permission, C. T. Sari and C. Gunduz-Demir, "Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images," in IEEE Transactions on Medical Imaging, vol. 38, no. 5, pp. 1139-1149, 2019.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the current practice of medicine, histopathological examination is the gold standard for diagnosing and grading many neoplastic diseases including cancer. This procedure requires a pathologist, who has extensive medical knowledge and training, to visually inspect a histopathological sample provided on a glass slide under a microscope (Figure 1.1(a)). However, as it mainly relies on the visual interpretation of the pathologist, this histopathological examination may become a complex and error-prone process, also depending on the complexity of the case. On the other hand, the augmentation of low-cost whole slide digital scanners provides digitized histopathology slides at high resolutions and the examination of these digitized images has begun to replace the traditional glass slide examination process (Figure 1.1(b)). Digital pathology targets at presenting various computerized tools and methods to diagnose plenty of diseases by analyzing these digitized histopathology slides in a fast and objective manner.

In the digital pathology literature, traditional approaches aim at alleviating these problems by providing computerized methods that quantitatively introduce handcrafted features. The performance of these methods mainly relies on the features that they use, and thus, their success strictly depends on the ability of these features by successfully quantifying the histopathology domain. Deep architectures have been introduced to overcome the feature extraction problems

(a)

(b)

Figure 1.1: (a) Routinely used histopathological examination process. (b) Histopathological examination in a digital pathology system.[1]

of traditional methods and have provided accurate results for classification and segmentation tasks. On the other hand, most of the existing deep learning based methods do not employ unsupervised learning at all or use them in a limited and inadequate way, thus they are unable to benefit from the unannotated data that cover the majority of the data available. This thesis aims to obviate the limitations of traditional and deep learning approaches and proposes two deep learning methods for classification and segmentation of histopathological images.

[1]Illustrations are taken from the following links, respectively: `https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/surgical-pathology`, `https://proscia.com/company/what-is-digital-pathology/`.

1.1 Motivation

Inspection and diagnosis of colorectal cancer are traditionally conducted by pathologists' manual examination of histopathological glass slide samples taken from patients. Although this traditional method has been used successfully for many years to make accurate diagnoses, it has several disadvantages [1, 2, 3]. Since glass slide samples are physical assets, they can be found in a single geographic location at a given time, which imposes the ability to view and inspect samples only by a group of pathologists present at the given location. Additionally, these glass slide samples should be maintained under appropriate conditions, thereby preventing them from deforming, deteriorating, and getting lost over time. The precautions and procedures to be considered to meet these conditions are not only costly but also very risky in terms of adversely affecting the samples in time. Another limitation of working on physical samples is that it is difficult to query a sample with certain filter parameters (e.g., patient name, sample date, and sample type) and obtain it on demand. Lastly, only a single glass slide sample can be examined at a time with microscopes used in traditional methods, which makes it troublesome to examine different samples together and also to swiftly examine other areas adjacent to the sample area by aligning them accurately.

High technology whole slide digital scanners developed in recent years enable to relocate the histopathological samples from glass slides to digital environment rapidly and reliably with low costs. Digital pathology aims to find fast and robust solutions for histopathological image analysis tasks with the algorithms and methods designed and developed on the digitized images produced by these scanners. There are many advantages of using and analyzing digital images produced with these scanners by digital pathology methods instead of storing and examining traditional glass slide samples manually [1, 2, 3]. First of all, storing histopathological samples in digital media and cloud environments instead of physical warehouses enables the samples to be examined by more than one medical institution simultaneously so that multiple diagnostics on a case can be obtained quickly. Secondly, with inexpensive storage and backup solutions, the

quality of the samples taken from a patient at the time of collection is ensured to be protected without changing over time. Thirdly, digital images can be easily queried with the desired parameters thanks to the digital pathology software provided. Fourthly, it is possible to examine the samples taken at different resolutions and magnification levels concomitantly with the software and hardware tools proposed by digital pathology. Additionally, these tools align the adjacent samples accurately, enabling them to navigate through these samples and establish a wider perspective for diagnosis. Last but not least, digital pathology contributes to the accuracy of the diagnosis of pathologists with traditional and new generation machine learning approaches and even starts to replace manual diagnoses in easily distinguishable samples, which comprise the majority of the cases [4].

Up to recent studies, the machine learning methods developed for digital pathology typically rely on defining and extracting handcrafted quantitative features from histopathological images and using these features in the design of a classification or a segmentation algorithm. These traditional classification and segmentation methods yield promising results in numerous digital pathology applications. On the other hand, the features that they use are handcrafted and strictly dependent on medical expert knowledge. Quantitatively expressing medical expert knowledge might be quite difficult for some applications, and thus, learning features directly from image data has the potential to generate features that represent the images better. In order to define more expressive and more robust features, deep learning based studies have been proposed to learn the features directly on image data without the need for handcrafted feature extraction procedures. For that, the majority of these studies train a supervised convolutional neural network (CNN) model [5] and exploit its output for classification or segmentation purposes.

Although supervised deep architectures lead to promising results regardless of the type of the computer vision task, they have a major limitation. Since the majority of deep approaches employing CNN architectures are trained in a supervised manner, they require annotated images for both classification and segmentation purposes. The augmentation in the number of pathology cases produces a vast

amount of digitized images that need to be examined, and pathologists are inadequate for the annotation procedures that should be done to use these images in a supervised learning method. Even if pathologists can perform annotation procedures to some extent, this may not be sufficient for accurate classification or segmentation method. Concretely speaking for insufficiency of annotated datasets, many of the studies crop small patches out of the images, train a learner on the patches, and then use the patch labels for entire image classification or segmentation. For that, they label a patch with the type of the segmented region covering this patch if the focus is segmentation. Otherwise, if it is classification, they label a patch with the class of the entire image without paying attention to the local characteristics of its subregions since the latter type of labeling is quite difficult and extremely time-consuming. On the other hand, considering the local characteristics of patches/subregions in a classifier may improve the performance since a tissue contains subregions showing different local characteristics and the distribution of these subregions determines the characteristics of the entire tissue.

The aforementioned inadequacy of annotated data leads researchers to design methods that incorporate unsupervised learning into supervised learning to benefit from unannotated data for both classification and segmentation purposes. Unsupervised learning aims to produce effective solutions using different approaches for histopathological image analysis tasks. In the feature extraction approach, which is one of the most frequently used approaches, the proposed methods have aimed at reconstructing images that feed the input layer, in the output layer, thus obtaining higher representations of the input images without the need for any ground truth data. There are studies using unsupervised learning for feature extraction in the histopathological image analysis literature, but they have some deficiencies in the context of this thesis. The models proposed in some of these studies aim to either classify [6] or detect [7] cytological elements (e.g., nucleus) within whole slide histopathological images by concentrating on small patches and do not provide any solution for the classification or segmentation of larger histopathological images. In [8], the proposed method obtains features from local patches using unsupervised learning and exploits these features to classify the whole image. However, the method is trained using all the

patches within the image, regardless of whether they are relevant or prominent, and without including any prior information in the model. Although there are methods to identify region proposals within the image and train their models on these regions [9, 10], these methods also have limitations. Since these methods are introduced to be used on natural images, the networks proposed in these methods have also been pretrained on a general-purpose dataset. However, this pretraining may not contribute to the fine-tuning step which would be held on a histopathological dataset. In addition to this, although these methods identify region proposals by using unsupervised learning, they train a supervised CNN to obtain higher representations of these proposals and therefore need ground truth data for the given input.

In addition to using it in the feature extraction step, there are other studies employing unsupervised learning as a regularization tool to improve the performance of supervised learning. Earlier studies of this approach have used layer-wise unsupervised pretraining to initialize weights, which are then finetuned by supervised training using backpropagation. This pretraining may provide regularization on backpropagation by enabling it to start with a better solution and may improve the network's generalization ability [11]. On the other hand, it has been argued that the weights learned by pretraining may be easily overwritten during supervised training [12] or even they may not provide a better initial solution at all [13] since the network is pretrained independently and by being unaware of the supervised task.

For more effective regularization, recent studies have trained a multi-task network to simultaneously minimize supervised and unsupervised losses by backpropagation [12, 13, 14, 15]. They define the supervised loss on the main classification task and the unsupervised loss on an auxiliary image reconstruction task. These two tasks typically share an encoder path to extract feature maps, from which a decoder path reconstructs an image and a classification path estimates a one-hot class label. In [15], in addition to this, another autoencoder with its own encoder and decoder is used and the outputs of the two decoders are combined to reconstruct the image. These studies calculate the reconstruction loss between original and decoded images as well as between the maps of the corresponding

intermediate layers of the encoder and decoder. In [13], noisy original images are used as inputs and the reconstruction loss is calculated between these images and their denoised versions.

All these studies define losses on the classification and reconstruction tasks separately and linearly combine them in a joint loss function, which they use to simultaneously learn these two tasks. This may provide regularization since the tasks compete during backpropagation. On the other hand, the effectiveness of this regularization highly depends on to what extent the supervised and unsupervised losses contribute to the joint loss function. When the unsupervised loss contributes too much, the network may not sufficiently learn the main classification task. When it contributes too small, the network may not learn the auxiliary reconstruction task, which results in not getting the expected regularization effect from unsupervised learning. Thus, these studies necessitate externally selecting right contributions that yield balanced learning between the supervised and unsupervised tasks. However, depending on the application, this external selection may not be always straightforward. It may become even harder when the joint loss includes more than one reconstruction loss (e.g., the one at the input level and those at the intermediate layers).

1.2 Contribution

Unsupervised learning has been exploited in various types of deep architectures for numerous medical image analysis tasks [16, 17]. Many of these methods mainly employ unsupervised learning for two purposes, either to extract features from data without the need of having the ground truth or to regularize a supervised learner to improve its classification/segmentation performance. 1) As a feature extractor, an unsupervised learning task aims to obtain higher-level representations learned directly from image data. To this end, it basically introduces a deep network that consists of a set of consecutive architectures (e.g., autoencoders [18, 19], restricted Boltzmann machines (RBMs) [20]) as hidden layers and this deep network is trained to reconstruct the image data itself at the

output layer. Since each hidden layer within the deep network presents a higher representation, the output of one or more intermediate layers are obtained and used as the feature set of the image in a supervised learner. 2) As a regularization tool, the reconstruction of image data is considered as an auxiliary unsupervised task and incorporated into the training procedure simultaneous to the supervised task. The unsupervised image reconstruction task is considered to be strongly related to the main supervised task since learning the input distribution simultaneously can contribute to the learning of the supervised task [11]. Weights shared by unsupervised and supervised networks are trained simultaneously by employing a joint function of unsupervised and supervised losses and the auxiliary unsupervised task aims to improve the performance of the main supervised task.

This thesis addresses the issues mentioned in the previous subsection by providing new solutions that incorporate unsupervised learning into classification and segmentation methods for histopathological image analysis in terms of both feature extraction and training regularization purposes. Thereby, it introduces two deep learning methods for the purpose of classification and segmentation of histopathological images.

The first study of this thesis proposes a novel semi-supervised method for the classification of histopathological colon tissue images [21]. In this context, the study has two main contributions. As the first contribution, it proposes to benefit from prior domain-knowledge provided by the pathologists' insight and expertise. A tissue is visually characterized by the traits of its cytological components, which are determined by the appearance of the components themselves and the subregions in their close proximities. In a typical examination procedure of a histopathological tissue image, pathologists visually inspect a tissue sample by focusing on salient regions located around the important sections of the tissue. They diagnose and grade cancer via examining the close proximities of the cytological components instead of focusing randomly selected subregions. Inspired by this, this study proposes to characterize the tissue image by first identifying its salient subregions and then using only these subregions for the training of the

deep architecture. There are existing deep learning approaches for histopathological image analysis that crop small patches out of images, train a learner on the patches, and then use the patch labels for the entire image classification [22, 23], nucleus detection [6, 24, 25], or entire image segmentation [26]. As opposed to our proposed method, these studies either pick random points in an image as the patch centers, or divide the image into a grid, or use the sliding window approach. None of them identify salient subregions/components and use them to determine their patches.

As the second contribution, the study devises an unsupervised method for the characterization of the salient subregions. With this proposed method, it was aimed to benefit from the effectiveness of unsupervised learning in feature extraction. The method pretrains a deep belief network, consisting of consecutive RBMs, on these salient subregions, allowing the system to extract high-level features directly from image data. To do so, this unsupervised feature extractor proposes to use the activation values of the hidden unit nodes in the final RBM of the pretrained deep belief network and to feed them into a clustering algorithm for quantizing the salient subregions (their corresponding cytological components) in an unsupervised way. The characterization of salient subregions using an unsupervised feature extractor allows us to obtain features without the need for expensive and impractical annotating of images. Although there exist annotated histopathological image datasets, the annotations in these datasets are usually at the entire image level. This causes all the patches extracted from the image to be labeled with the entire image class and the models to be trained on these patches are not able to encapsulate the local characteristics within the image. The proposed unsupervised feature extractor prevents this problem and enables a more robust and expressive training. To the best of our knowledge, this study is the first example that successfully uses a deep belief network of RBMs for the characterization of histopathological tissue images.

In order to benefit from the effectiveness of unsupervised learning in regularization, **the second study** proposes an effective method to combine the supervised and unsupervised tasks to train a fully convolutional network for the task of semantic segmentation in histopathological images. This solution relies

on defining a new embedding that unites the main task of segmentation and an auxiliary task of image reconstruction into a single task and learning this united task by a single generative model. To this end, it first introduces an embedding that generates a multi-channel output image, on which segmentation is trivial, by superimposing an input image on its segmentation map. Then, it proposes to learn this newly generated output image from the input image using a conditional generative adversarial network (cGAN), which is known to be effective for image-to-image translations. This new embedding together with its learning by a cGAN provide two main contributions. As the first contribution, the proposed embedding unites segmentation and reconstruction tasks, which concomitantly results in combining supervised and unsupervised objectives (losses) in a very natural way. This presents an alternative to externally determining the contributions of these tasks in a joint loss function. More importantly, since the output image of the united task corresponds to a segmentation map that preserves a reconstructive ability, uniting the segmentation and reconstruction tasks enforces the network to jointly learn image features and context features. This joint learning provides effective regularization. This training regularization is obtained since reconstructing the input image, and hence, capturing the input image distribution $P(X)$ contributes to the learning of the segmentation task $P(Y|X)$ [11]. In addition to this, learning these two tasks simultaneously prevents unsupervised task from learning trivial representations that do not contribute to supervised task [12]. As the second contribution, the proposed method learns the output image of the united task by benefiting from the well-known synthesizing ability of cGANs. Thanks to using a cGAN, the method produces more realistic outputs that adhere to spatial contiguity without any post-processing (e.g., using conditional random fields, CRFs [27]). To the best of our knowledge, this is the first proposal of using a cGAN to produce such embedded output images that can be directly used for semantic segmentation.

1.3 Outline

The remainder of this thesis is organized as follows. The medical background for histopathological images used in the proposed studies and the related literature in the context of histopathological image analysis are presented in Chapter 2. A novel semi-supervised method for the classification of histopathological colon tissue images is deeply discussed in Chapter 3. In Chapter 4, an effective semantic segmentation method for various types of histopathological images is described in detail. Finally, the summary of this thesis and the future research directions are given in Chapter 5.

Chapter 2

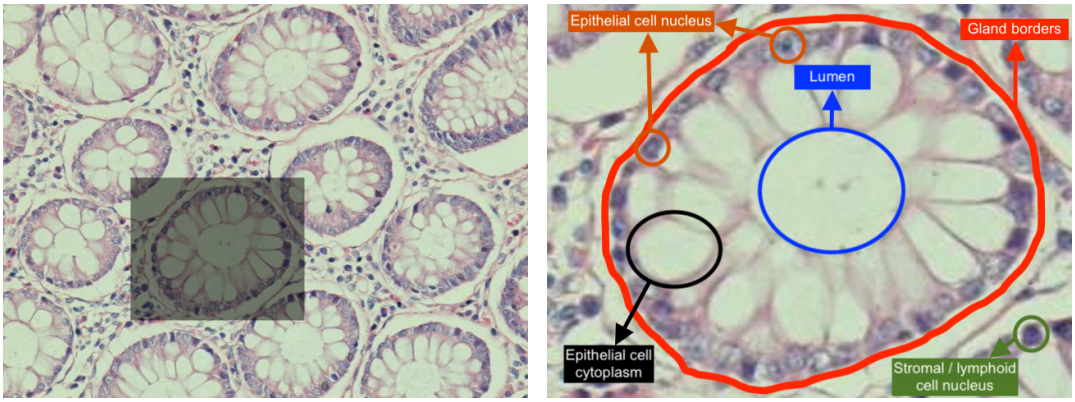
Background

This thesis proposes deep learning approaches for the classification and segmentation of histopathological images of colon tissues. The medical background related to this thesis work is briefly explained in Section 2.1. The related literature in the context of traditional and recent deep learning methods for classification and segmentation of histopathological images is provided in Section 2.2.

2.1 Medical Background

One out of every six deaths in the world is caused by cancer, which places it in the second leading rank among all diseases that cause death [28]. Colorectal cancer is the third most common cancer type and is placed in the fourth rank among all cancer-caused deaths [29]. Colon adenocarcinoma is the most common form of colorectal cancer, accounting for about 90 percent of cancer cases in North America and Western Europe.

The diagnosis and grading of colon adenocarcinoma are conducted with the manual examination of histopathological tissues under a microscope. In a typical colon tissue, a lumen is located in the center of a gland with epithelial cell nuclei and cytoplasms lined up around it (Figure 2.1). These epithelial cells and luminal



(a) (b)

Figure 2.1: (a) Glands in a normal colon tissue sample. (b) A colon tissue consists of cytological components including cellular, stromal, and luminal components.

components form glandular structures. Colon adenocarcinoma is originated from epithelial cells and leads to distortions and disaggregations on these cells as well as on the glands, which are formed by these epithelial cells. Glandular structures in a normal colon tissue are illustrated in Figures 2.2(a) and 2.2(b). When cancer occurs at the initial level, relatively low distortions begin to appear within colon tissues and glandular structures and formation of these tissues are well to moderately differentiated (Figures 2.2(c) and 2.2(d)). In such a sample of colon tissue, glands can still be differentiated, but the boundaries of these structures begin to lose their clarity. With the progression of cancer, the level of distortion increases and glandular structures of these tissues are only poorly differentiated or may not be differentiated at all (Figures 2.2(e) and 2.2(f)).

The scope of this thesis covers the classification and segmentation of histopathological images of colon tissues. To this end, the first study of the thesis proposes a semi-supervised method for the classification of homogeneous colon tissue images, in which each image sample covers a part of a tissue belonging to a single class. Since stroma is the supporting material of colon tissues, normal, low-grade, and high-grade cancerous samples contain stroma among their glandular structures, and stroma is not considered as a distinct separate class. Examples of normal, low-grade, and high-grade adenocarcinomatous (cancerous) colon tissue images are given in Figure 2.2 and more details about this dataset

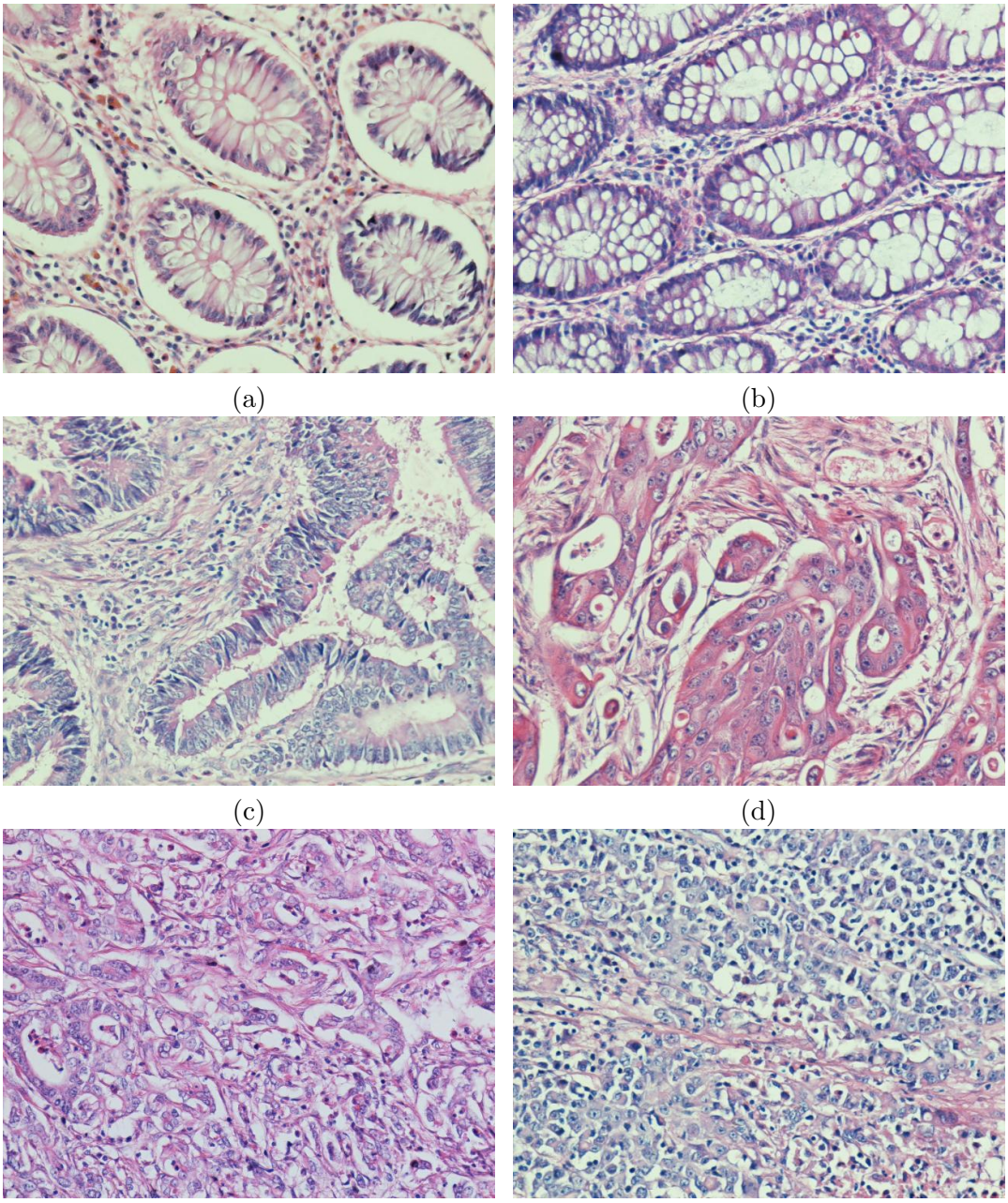


Figure 2.2: Images of homogeneous colon tissues classified as (a)-(b) normal, (c)-(d) low-grade cancerous (adenocarcinomatous), and (e)-(f) high-grade cancerous (adenocarcinomatous).

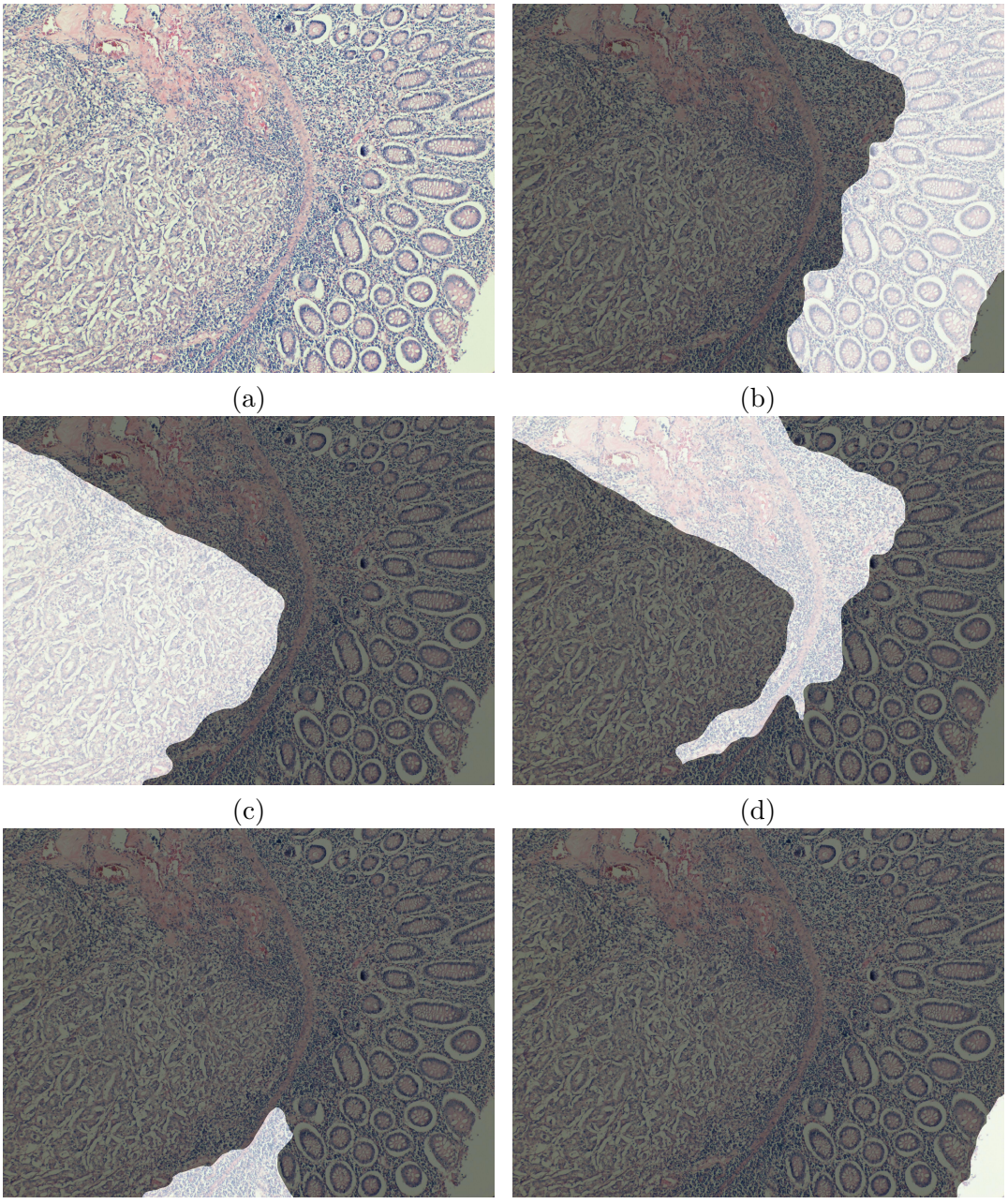


Figure 2.3: (a) A sample heterogeneous colon tissue image. Highlighted regions are annotated as (b) normal, (c) tumorous, (d) connective tissue, (e) dense lymphoid tissue, and (f) empty.

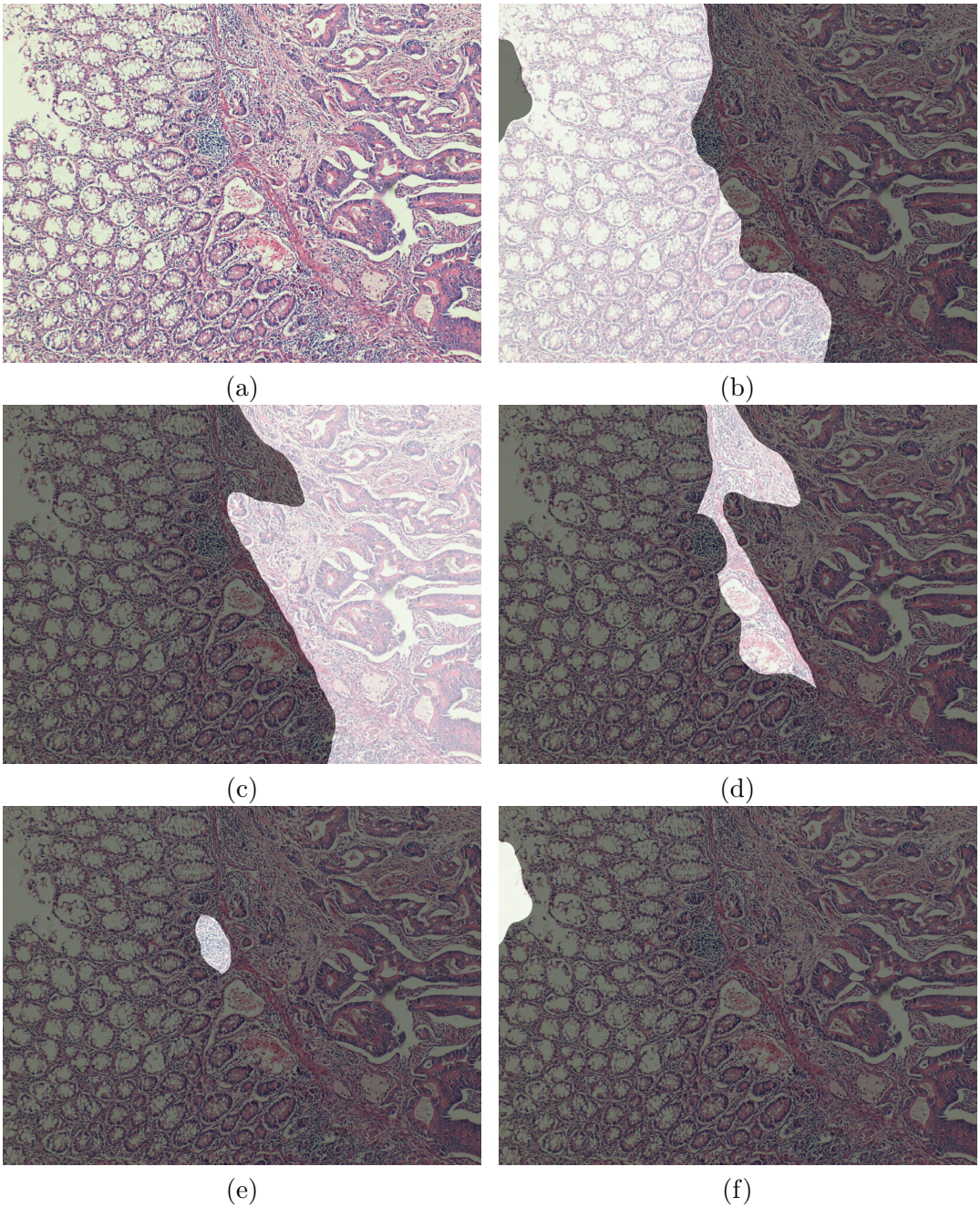


Figure 2.4: (a) A sample heterogeneous colon tissue image. Highlighted regions are annotated as (b) normal, (c) tumorous, (d) connective tissue, (e) dense lymphoid tissue, and (f) empty.

are to be given in Chapter 3.

The second study of the thesis proposes a semantic segmentation method for relatively larger colon tissue images. Typically, larger colon tissue images are heterogeneous since each image contains non-overlapping regions of differently annotated classes. Figures 2.3(a) and 2.4(a) illustrate sample heterogeneous colon tissue images and the regions of different classes within these images are annotated with the corresponding classes by an expert pathologist. In Figures 2.3(b) and 2.4(b), normal tissue regions are annotated. Although the regions in Figures 2.3(c) and 2.4(c) contain different levels of distortions, and therefore, are of different cancer grades, both regions are annotated as tumorous (cancerous) in the context of this second study. The regions containing stroma, which is the connective material between glands, are annotated as connective tissue and shown in Figures 2.3(d) and 2.4(d). The lymphoid aggregates are annotated as dense lymphoid tissue and shown in Figures 2.3(e) and 2.4(e). The empty glass and debris regions are annotated as empty and shown in Figures 2.3(f) and 2.4(f). More details about this dataset are to be given in Chapter 4.

2.2 Related Work

Digital pathology has been introduced to provide auxiliary tools for manual examinations of histopathological samples conducted by pathologists and to diagnose patients more accurately and objectively using automated or semi-automated methods. To this end, the proposed methods aim at finding solutions to the classification and segmentation problems that pathologists study on histopathological samples. Earlier studies in the digital pathology literature have proposed to extract handcrafted features from histopathological images for their representation. These studies mainly rely on two feature extraction approaches. The textural approach quantifies the spatial arrangement of pixel intensities and defines the textural features using intensity histograms [30, 31, 32], co-occurrence

matrices [33, 34], wavelets [35], fractal analysis [36, 37], and local binary patterns [38, 39, 40]. In the structural approach, features are obtained by characterizing the spatial distribution of cytological components within the image. Most of the studies form graph representations by considering cytological components as nodes and use these graph representations to calculate the feature set of the given image [41, 42, 43]. Earlier studies of our research group have employed nuclear, stromal, and luminal tissue components as nodes of a graph representation and quantified their spatial representation using the graph representation [44, 45, 46]. Although both textural and structural approaches perform well in numerous histopathological image applications, defining expressive handcrafted features may require significant insight on the corresponding application. However, this is not always that trivial and improper feature definitions may greatly lower algorithms' performance.

Recently, deep learning methods have shown great promise for various computer vision tasks, and this has led researchers to use deep learning methods in histopathological image classification and segmentation. Besides, the fact that deep learning methods learn features directly from data and do not need any external support enables their use in histopathological image analysis. Similar to many computer vision tasks, the most preferred methods in histopathological image analysis are CNN models and their variations. In the histopathological domain, these models are generally trained in a supervised manner and the outputs produced by these models are used for classification and segmentation. Most of these methods use one of the two approaches. In the first one, methods train a CNN on entire histopathological training images, feed an entire histopathological test image to the trained CNN, and use the class label it outputs to directly classify the test image [47, 48, 49]. In the second approach, methods divide each histopathological image into a grid of patches, feed each test patch to the CNN, which is trained on the same-sized training patches, and then exploit either the class labels or the posteriors generated by the CNN. In [22], the labels are voted to classify the image out of which the patches are cropped. In [26], the patch labels are directly used to segment the tissue image into its epithelial and stromal regions. These patch labels are also employed to extract structural features,

which are then used for whole slide classification [50, 51] and gland segmentation [52]. In [53], a CNN model, which is pretrained on a general-purpose dataset, is employed to extract features from overlapping patches in histopathological tissue images. These features are then classified by a linear SVM and each pixel is classified by the majority voted class of the patches covering this pixel since the pixel is classified within several patches. The method introduced in [54] inserts an SE-ResNet [55] module between convolutional and fully connected layers of the proposed deep CNN architecture to reduce the number of parameters and prevent overfitting for the classification of breast cancer histology images. Although they are not histopathological images, this patch-based CNN approach is used to differentiate nuclear and background regions in fluorescence microscopy images [56] and nuclear, cytoplasmic, and background regions in cervical images [57].

The posteriors generated by a supervised CNN are commonly used to segment a tissue image into its regions of interest (ROI). To this end, for the class corresponding to the ROI (e.g., nucleus or gland class), a probability map is formed using the patch posteriors. Then, the ROI is segmented by either finding local maxima on this posterior map [24, 25, 58, 59] or thresholding [60]. This type of approach has also been used to detect cell locations in different types of microscopic images such as live cell [61], fluorescent [62], and zebrafish [63] images. As an alternative, nuclei are located by post-processing the class labels with techniques such as morphological operations [64] and region growing [65]. In [66], after obtaining a nucleus label map, nuclei's bounding boxes are estimated by training another deep neural network. In a more recent study [67], authors propose a multi-stage network in which a patch-level CNN is trained on a histopathological multi-organ dataset to generate pixel-level activation maps for independent patches within images and inter-patch adjacencies are incorporated by applying mathematical operations, averaging, and post-processing (with a CRF) to obtain final segmentation maps. Another multi-stage network [68] combines a patch-based classification model with a whole slide-scale segmentation model for whole slide image (WSI) segmentation. For that, patches cropped from WSIs are first used to train the patch classifier and the output of an intermediate layer is used as feature vectors of the patches. Then, the patch features are arranged based on

their position on the WSI and the segmentation model is trained on whole slide feature maps obtained by the local patch features to produce the final whole slide prediction map.

Although CNN-based models considering local adjacencies are very successful for histopathological image analysis tasks, particularly for classification problems, they also have some disadvantages. Since histopathological images generally have very large dimensions, CNN-based models applied to entire images suffer from an excessive number of parameters. On the other hand, the patch-based CNN models, which are proposed to overcome this drawback, are also exposed to difficulties in determining the patch size and incorporating inter-patch neighborhood information. To overcome these challenges, fully convolutional networks (FCNs) have been proposed to provide efficient solutions for semantic segmentation [69]. The UNet architecture [70] proposed for biomedical image segmentation has become the state-of-the-art FCN model for semantic segmentation tasks in many fields of computer vision and has been frequently preferred to predict pixel-level class labels for histopathological image segmentation [71, 72]. It has been also proposed to fuse the predictions of multiple FCNs. In [73], FCNs are trained on images of different resolutions. In [74], they are constructed by starting the upsampling operation from different layers of the same encoder. Other studies perform segmentation at finer-levels; they usually segment nucleus and gland instances. They typically use multi-task networks, in which auxiliary tasks are defined as predicting boundary of instances [75] and their bounding boxes [76]. Application specific additional tasks, such as lumen prediction [77] and malignancy classification [78], are also used for gland instance segmentation. Note that the focus of this thesis is compartment segmentation at the tissue level but not instance segmentation.

FCNs are typically trained to predict pixel labels independent of each other. This may prevent to capture local and global spatial contiguity within an entire image. To recover fine details, CRFs using pair-wise potentials have been employed as a post-processing step to refine the segmentation maps generated by FCNs [27, 79]. Although CRFs lead to improvements, the integration of FCNs and CRFs with higher orders is limited [80] and using such additional layers,

which are externally added to the end of FCNs and are not trained with FCNs simultaneously, breaks the end-to-end architecture of models.

Unsupervised learning is exploited in many different fields of medical image analysis to improve supervised learning models that perform classification and segmentation. There exist studies that make use of unsupervised learning in their systems to extract features to be used in a classification or a segmentation method [81]. In [8], a set of autoencoders are first pretrained on small image patches and the weights of each autoencoder are employed to define a filter for the first convolution layer of a supervised CNN classifier, which is then to be used to classify an entire tissue image. Similarly, in [7], a stacked autoencoder is pretrained on image patches and the outputs of its final layer are fed to a supervised classifier for nucleus detection. As opposed to our first study, these previous studies did not cluster the outputs of the autoencoders to label the patches in an unsupervised way and did not use the label distribution for image classification. The study in [23] is similar to our first study in the sense that it also clusters the patches based on the outputs of a stacked autoencoder. However, this study did select its patches randomly and did not consider any saliency in a tissue image. On the contrary, our first study proposes to determine the salient subregions by prior domain-knowledge, characterize them by an unsupervised deep belief network consisting of consecutive RBMs, and use the characteristics of *only* these salient subregions to classify the entire tissue image. Our experiments have demonstrated that the use of saliency together with this unsupervised characterization improve the accuracy. Additionally, as opposed to all these previous studies, which employ either a CNN or a stacked autoencoder, our study uses a deep belief network of restricted Boltzmann machines.

In addition to using unsupervised learning to extract features from image data, many studies in the literature exploit it to regularize the supervised training of classification or segmentation networks. To regularize the supervised training, the earlier studies have used multi-task networks that consider complementary tasks along with the main task of segmentation. These are the networks with a shared encoder and parallel decoders, one for each task, and they are trained to minimize the joint loss defined on all decoders [75].

Another way of regularization is to use unsupervised learning in the form of defining an additional image reconstruction task and learning it concurrently with the main task. Most of the previous studies focus on non-dense prediction, defining their main task as to predict one-hot class label for an entire image [12, 13, 14, 15]. Only a few consider the main task of image segmentation [82, 83]. However, all these studies use image reconstruction as an auxiliary task and linearly combine its loss and the loss of classification/segmentation, which are defined independently, in a joint loss function. This is different than our second study, which unites the image reconstruction and segmentation tasks through its proposed embedding and trains its network to minimize the loss on this united task. Moreover as opposed to our second study, these previous studies do not use a generative adversarial network for their network.

Aforementioned limitations of CNNs, single-task and multi-task FCNs lead researchers to use or design new architectures for histopathological image classification and segmentation. Generative adversarial networks (GANs) are firstly proposed for image synthesis by using two networks, generator and discriminator, trained in an adversarial manner. The first applications of GANs in the field of histopathological image analysis are also for data synthesis purposes [84, 85, 86] since the amount and variety of data in histopathology domain is insufficient. Meanwhile, GANs are also exploited to extract features from histopathological images and train a classifier on these features. In [87], a unified GAN architecture is employed to learn and extract cell-level features in histopathology images and these features are used for image-level classification. Its application to semantic segmentation typically provides an additional input to the generator (segmentor) to control its output [88, 89]. Adversarial loss has also been used to regularize network training. One work [90] uses it for an autoencoder to better learn its feature maps. It considers the encoder as the generator and feeds its outputs to the discriminator. Then, it updates encoder weights considering the adversarial loss in addition to the reconstruction loss between encoder's input and decoder's output. Another work [91] estimates a segmentation map from an image and then reconstructs the image from the estimated map for regularization. It uses a cGAN for image reconstruction, and hence, employs the adversarial loss in addition to

the segmentation and image reconstruction losses. However, it also separately defines these losses and linearly combines them in a joint loss function. None of these previous studies exploit an embedding to combine supervised and unsupervised losses for regularizing their network for semantic segmentation. Different than our second study, none of these studies define an embedding to unite the segmentation and image reconstruction tasks and use a cGAN to learn this united task. Only a few use a cGAN for nucleus and gland segmentation [86, 92]. However, these studies define adversarial loss on the genuineness of their segmentation maps but they do not consider image reconstruction loss in their segmentation networks. Besides, they do not use any embedding to regularize training.

Chapter 3

Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images

3.1 Methodology

Our proposed method relies on representing and classifying a tissue image with a set of features extracted by a newly proposed unsupervised feature extractor. This extractor defines the features by quantifying only the characteristics of the salient subregions in the image instead of considering those of all image locations. To this end, it first proposes to define the salient subregions around cytological tissue components (Section 3.1.1). Afterwards, to characterize the subregions/components in an unsupervised way, it learns their local features by a deep belief network consisting of consecutive RBMs and quantizes them by clustering the local features by the k-means algorithm (Section 3.1.2). At the end, it represents and classifies the image with the distribution of its quantized subregions/components (Section 3.1.3). A schematic overview of the proposed

method is given in Figure 3.1 and the details of its steps are explained in the following subsections. The source codes of its implementation are available at http://www.cs.bilkent.edu.tr/~gunduz/downloads/DeepFeature.

The motivation behind this proposed method is the following: A tissue contains different types of cells that serve different functions in the tissue. The visual appearance of a cell and its surrounding may look differently depending on the cell's type and function. Furthermore, some types of cells may form specialized structures in the tissue. The tissue is visually characterized by the traits of all these cytological components. Depending on its type, cancer causes changes in the appearance and distribution of certain cytological tissue components. For example, in colon, epithelial cells line up around a lumen to form a gland structure and different types of connective tissue cells in between the glands support epithelia. In a normal tissue, the epithelial cells are arranged in a single layer and since they are rich in mucin, their cytoplasms appear in light color. With the development of colon adenocarcinoma, this single layer structure is getting disappeared, which causes the epithelial cells' nuclei to be seen as nucleus clutters, and their cytoplasms return to pink as they become poor in mucin. With the further progression of this cancer, the epithelial cells are dispersed in the connective tissue and the regular structure of a gland gets totally lost (see Figure 3.2). Some of such visual observations are easy to express, but some others may lack of a clear definition although they are in the eyes of a pathologist. Furthermore, when there exists a clear definition for an observation, its expression and quantification commonly require exact component localization, which emerges a very difficult segmentation problem even for a human eye, and its use in a supervised classifier requires very laborious annotation. Thus, our method approximately represents the tissue components with a set of multi-typed circular objects, defines the local windows cropped around these objects as the salient subregions, and characterizes them in an unsupervised way. Note that this is just an approximate representation where one object can correspond to multiple components or vice versa. It is also worth to noting that the salient subregions cropped around the objects are defined with the aim of approximately representing the components, whose characterizations will further be used in the entire image characterization.

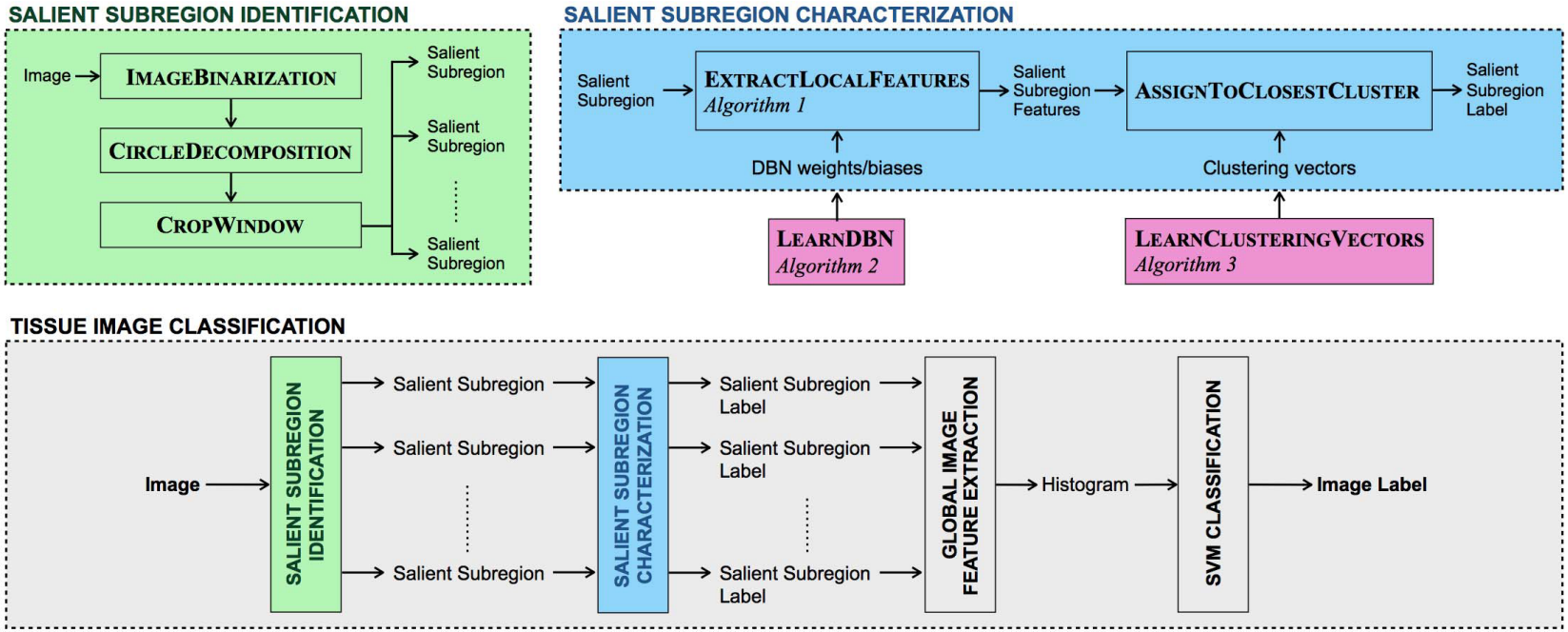


Figure 3.1: A schematic overview of the proposed method.

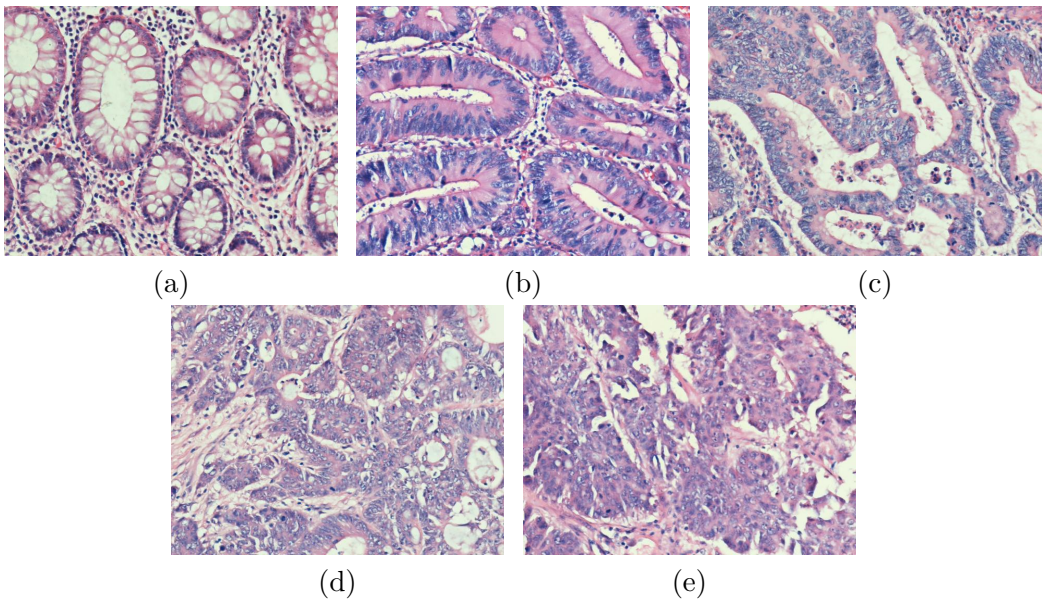
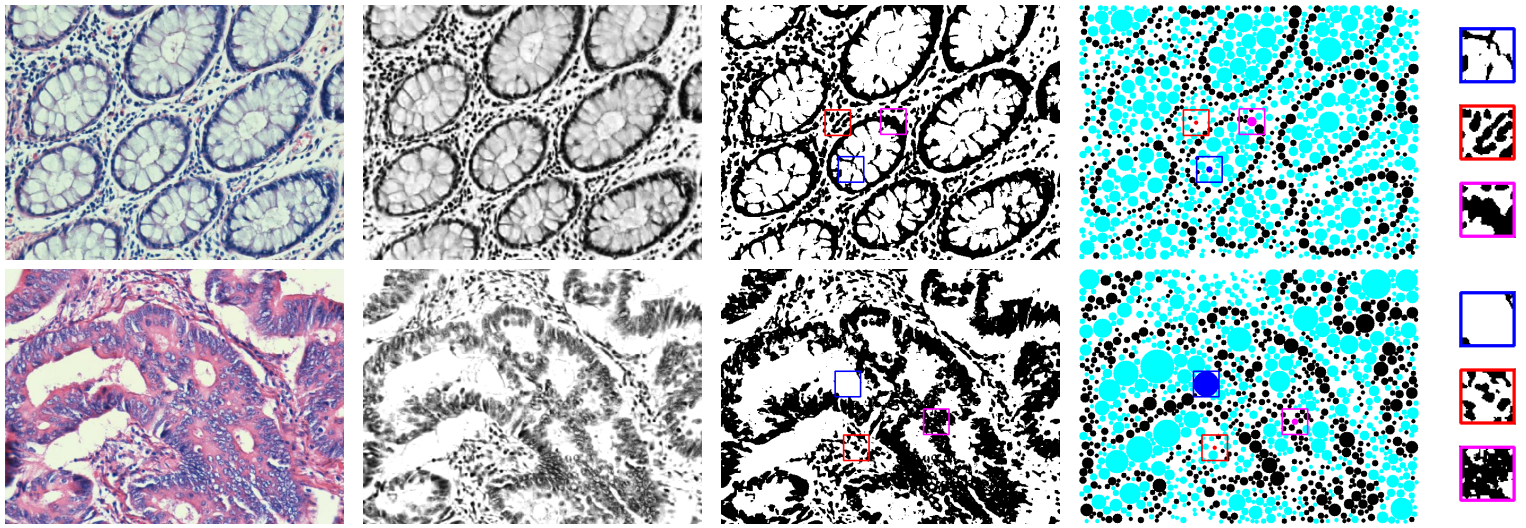


Figure 3.2: Example images of tissues labeled with different classes: (a) Normal, (b) low grade cancerous--*grade1*, (c) low grade cancerous--*at the boundary between grade1 and grade2*, (d) low grade cancerous--*grade2*, and (e) high grade cancerous. Note that the normal and high grade cancerous classes are the same for our first and second datasets whereas the low grade cancerous class in the first dataset is further categorized into three in the second one.

3.1.1 Salient Subregion Identification

Salient subregions are defined around tissue components whose locations are approximated by the algorithm that we previously developed in our research group [46]. This approximation and salient subregion identification are illustrated on example images in Figure 3.3 and the details are explained below.

The approximation algorithm uses nuclear and non-nuclear types for object representation. For that, it first separates the hematoxylin channel of an image $\mathcal{I}$ by applying color deconvolution [94] and thresholds this channel to obtain the binary image $\mathcal{BW}$. In this thresholding, an average is calculated on all pixel values and a pixel is labeled as nucleus if its value is less than this threshold and non-nucleus otherwise. Then, the circle-fit algorithm [93] is applied on the pixels of each group in $\mathcal{BW}$ separately to locate a set of nuclear and non-nuclear objects. The circle-fit algorithm iteratively locates non-overlapping circles on the



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

Figure 3.3: (a) Original images; top is normal and bottom is cancerous, (b) hematoxylin channels obtained by stain deconvolution, (c) binary images obtained by thresholding, (d) circular objects located by the circle-fit algorithm [93], and (e) examples of salient subregions cropped around the three example located objects. In (d), black and cyan circles represent nuclear and non-nuclear objects, respectively. In (c) and (d), the blue, red, and magenta squares indicate example salient subregions cropped around three example objects, which are also shown in blue, red, and magenta in (c). As seen in the examples given in (e), local properties of small subregions in an image of different types might be similar or different. On the other hand, the distribution of the local properties is different for different types of images.

given pixels, starting from the largest one as long as the radii of the circles are greater than the threshold r_{min}. At the end, around each object c_i, a salient region Ω_i is defined by cropping a window out of the binary image $\mathcal{BW}$ where the object centroid determines the window center and the parameter ω_{size} determines its size. Note that although the located objects are labeled with a nuclear or a non-nuclear type by the approximation algorithm, we just use the object centroids to define the salient regions, without using their types. Instead, we will re-type (re-characterize) these objects with the local features that will be learned by a deep belief network (Section 3.1.2).

The substeps of this salient subregion identification are herein referred to as ImageBinarization, CircleDecomposition, and CropWindow functions, respectively. We will also use these functions in the implementation of the succeeding steps. To improve the readability of the thesis, we provide a list of these functions and their uses in Table 3.1. Note that this table also includes other auxiliary functions, which will be used in the implementation of the succeeding steps.

3.1.2 Salient Subregion Characterization via Deep Learning

This step involves two learning systems: The first one, LearnDBN, acts as an unsupervised feature extractor for the salient subregions, and hence, for the objects that they correspond to. It learns the weights of a deep belief network of RBMs and uses the activation values of the hidden unit nodes in the final RBM to define the local deep features of the salient subregions. The second system, LearnClusteringVectors, learns the clustering vectors on the local deep features. This clustering will be used to quantize any salient subregion, which corresponds to re-typing the object for which this salient subregion is defined. The details of these learning systems are given in Section 3.1.2.1 and 3.1.2.2.

Table 3.1: Auxilary function definitions.

Function	**Definition**
$\mathcal{BW} \leftarrow$ ImageBinarization($\mathcal{I}$)	Binarizes the image $\mathcal{I}$ with respect to its hematoxylin channel (see Section 3.1.1).
$\mathcal{C} \leftarrow$ CircleDecomposition($\mathcal{BW}$, r_{min})	Locates a set $\mathcal{C}$ of circular objects on nuclear and non-nuclear pixels of the binary image $\mathcal{BW}$ with a minimum radius of r_{min} (see Section 3.1.1).
$\Omega_i \leftarrow$ CropWindow($\mathcal{BW}$, c_i, ω_{size})	Defines a salient subregion Ω_i by cropping a $w_{size} \times w_{size}$ window out of the binary image $\mathcal{BW}$ around an object c_i.
$[W, B] \leftarrow$ ContrastiveDivergence($\mathcal{D}_{\texttt{dbn}}, P$)	Pretrains a deep belief network on the dataset $\mathcal{D}_{\texttt{dbn}}$ of the salient subregions, by applying the contrastive divergence algorithm to each of its RBMs. The architecture of the deep belief network is denoted by the input parameter P. Returns the weight matrices W and the bias vectors B of the pretrained deep belief network.
$V \leftarrow$ KMeansClustering($\mathcal{D}_{\texttt{kmeans}}, K$)	Clusters the dataset $\mathcal{D}_{\texttt{kmeans}}$ of the local deep features of the salient subregions into K using the k-means algorithm and returns the clustering vectors V.
$l_i \leftarrow$ AssignToClosestCluster(ϕ_i, V)	Labels the salient subregion Ω_i with l_i, which is the id of the closest clustering vector in V, according to the local deep features ϕ_i of this salient subregion.

3.1.2.1 Deep Network Learning

The LearnDBN algorithm pretrains a deep belief network, which consists of consecutive RBMs. An RBM is an undirected graphical model consisting of a visible and a hidden layer and the symmetric weights in between them. The output of an RBM (the units in its hidden layer) can be considered as a higher representation of its input (the units of its visible layer). To get the representations at different abstraction levels, a set of RBMs are stacked consecutively by linking one RBM's output to the next RBM's input. In this work, the input of the first RBM is fed by the pixels of a salient subregion Ω_i, which is cropped out of the binary image $\mathcal{BW}$, and the output of the last RBM is used as the local feature set ϕ_i of this salient subregion; see Algorithm 1. In this algorithm, W_j and B_j are the weight matrix and the bias vector of the j-th RBM, respectively.

Algorithm 1 ExtractLocalFeatures

Input: salient subregion Ω_i, number H of RBMs in the pretrained deep belief network, weight matrices W and bias vectors B of the pretrained deep belief network

Output: local feature set ϕ_i of the salient subregion Ω_i

1: $\Pi_0 = \Omega_i$
2: **for** $j = 1$ to H **do**
3: $\quad \Pi_j = \text{sigmoid}(\Pi_{j-1}\, W_j \;+\; B_j)$
4: **end for**
5: $\phi_i = \Pi_H$

The LearnDBN function learns the weights and biases of the deep belief network by pretraining it layer by layer using the contrastive divergence algorithm [95]. For this purpose, it constructs a dataset $\mathcal{D}_{\texttt{dbn}}$ from randomly selected salient subregions of randomly selected training images. Algorithm 2 gives its pseudocode; see Table 3.1 for explanations of the auxiliary functions. Note that LearnDBN should also input the parameters that specify the architecture of the network, including the number of hidden layers (the number of RBMs) and the number of hidden units in each hidden layer.

Algorithm 2 LearnDBN

Input: training set $\mathcal{D}$ of original images, size ω_{size} of a salient subregion, minimum circle radius r_{min}, architecture P of the deep belief network
Output: weight matrices W and bias vectors B of the pretrained deep belief network

$\mathcal{D}_{\texttt{dbn}} = \emptyset$
for each randomly selected $\mathcal{I} \in \mathcal{D}$ **do**
  $\mathcal{BW} \leftarrow$ ImageBinarization($\mathcal{I}$)
  $\mathcal{C} \leftarrow$ CircleDecomposition($\mathcal{BW}$, r_{min})
  for each randomly selected $c_i \in \mathcal{C}$ **do**
    $\Omega_i \leftarrow$ CropWindow($\mathcal{BW}$, c_i, ω_{size})
    $\mathcal{D}_{\texttt{dbn}} = \mathcal{D}_{\texttt{dbn}} \cup \Omega_i$
  end for
end for
$[W, B] \leftarrow$ ContrastiveDivergence($\mathcal{D}_{\texttt{dbn}}, P$)

3.1.2.2 Cluster Learning

After learning the weights and biases of the deep belief network, the ExtractLocalFeatures function is used to define the local deep features of a given salient subregion. This work proposes to quantify the entire tissue image with the labels (characteristics) of its salient subregions. Thus, these continuous features are quantized into discrete labels. As discussed before, annotating each salient subregion is quite difficult, if not impossible, and hence, it is very hard to learn these labels in a supervised manner. Therefore, this work proposes to follow an unsupervised approach to learn this labeling process. To this end, it uses k-means clustering on the local deep features of the salient subregions. Note that the k-means algorithm learns the clustering vectors V on the training set $\mathcal{D}_{\texttt{kmeans}}$ that is formed up of the local deep features of randomly selected salient subregions of randomly selected training images. The pseudocode of LearnClusteringVectors is given in Algorithm 3. This algorithm outputs a set V of K clustering vectors. In the next step, an arbitrary salient subregion is labeled with the id of its closest clustering vector.

Algorithm 3 LearnClusteringVectors

Input: training set $\mathcal{D}$ of original images, size ω_{size} of a salient subregion, minimum circle radius r_{min}, number H of RBMs, weight matrices W and bias vectors B of the pretrained deep belief network, cluster number K
Output: clustering vectors V

$\mathcal{D}_{\texttt{kmeans}} = \emptyset$
for each randomly selected $\mathcal{I} \in \mathcal{D}$ **do**
  $\mathcal{BW} \leftarrow$ ImageBinarization($\mathcal{I}$)
  $\mathcal{C} \leftarrow$ CircleDecomposition($\mathcal{BW}$, r_{min})
  for each randomly selected $c_i \in \mathcal{C}$ **do**
    $\Omega_i \leftarrow$ CropWindow($\mathcal{BW}$, c_i, ω_{size})
    $\phi_i \leftarrow$ ExtractLocalFeatures(Ω_i, H, W, B)
    $\mathcal{D}_{\texttt{kmeans}} = \mathcal{D}_{\texttt{kmeans}} \cup \phi_i$
  end for
end for
$V \leftarrow$ KMeansClustering($\mathcal{D}_{\texttt{kmeans}}, K$)

3.1.3 Image Representation and Classification

In the last step, a set of global features are extracted to represent an arbitrary image $\mathcal{I}$. To this end, all salient subregions are identified within this image and their local deep features are extracted. Each salient subregion Ω_i is labeled with the id l_i of its closest clustering vector according to its deep features ϕ_i by the AssignToClosestCluster auxiliary function (see Table 3.1). Then, to represent the image $\mathcal{I}$, global features are extracted by calculating a histogram on the labels of all salient subregions in $\mathcal{I}$ (i.e., the characteristics of the components that these subregions correspond to). At the end, the image $\mathcal{I}$ is classified by a support vector machine (SVM) with a linear kernel based on its global features. Note that, this study uses the SVM implementation of [96], which employs the one-against-one strategy for multiclass classifications.

3.2 Experiments

3.2.1 Datasets

We test our proposed method on two datasets that contain microscopic images of colon tissues stained with the routinely used hematoxylin-and-eosin technique. The images of these tissues were taken using a Nikon Coolscope Digital Microscope with a 20× objective lens and the image resolution was 480 × 640. The first dataset is the one that we also used in our previous studies. In this dataset, each image is labeled with one of the three classes: normal, low-grade cancerous, and high-grade cancerous. It comprises 3236 images taken from 258 patients, which were randomly divided into two to form the training and test sets. The training set includes 1644 images (510 normal, 859 low-grade cancerous, and 275 high-grade cancerous) of 129 patients. The test set includes 1592 images (491 normal, 844 low-grade cancerous, and 257 high-grade cancerous) of the remaining patients. Note that the training and test sets are independent at the patient level; i.e., images taken from a slide(s) of a particular patient are used either in the training or the test set.

The second dataset includes a subset of the first one with the low-grade cancerous tissue images being further subcategorized. Here only a subset was selected since subcategorization was difficult for some images. Note that we also excluded some images from the normal and high-grade cancerous classes to obtain more balanced datasets. As a result, in this second dataset, each image is labeled with one of the five classes: normal, low-grade cancerous (grade1), low-grade cancerous (grade2), low-grade cancerous (at the boundary between grade1 and grade2), and high-grade cancerous. The training set includes 182 normal, 188 grade1 cancerous, 121 grade1-2 cancerous, 123 grade2 cancerous, and 177 high-grade cancerous tissue images. The test set includes 178 normal, 179 grade1 cancerous, 117 grade1-2 cancerous, 124 grade2 cancerous, and 185 high-grade cancerous tissue images. Example images from these datasets are given in Figure 3.2.

3.2.2 Parameter Setting

The proposed method has the following model parameters that should be externally set: minimum circle radius r_{min}, size of a salient subregion ω_{size}, and cluster number K. The parameters r_{min} and ω_{size} are in pixels. Additionally, the support vector machine classifier has the parameter C. In our experiments, the values of these parameters are selected using cross-validation on the training images of the first dataset without using any of its test samples. Moreover, this selection does not consider any performance metric obtained on the second dataset. By considering any combinations of the following values $r_{min} = \{3, 4, 5\}$, $\omega_{size} = \{19, 29, 39\}$, $K = \{500, 1000, 1500\}$, and $C = \{1, 5, 10, 25, 50, 100, 250, 500, 1000, 2500, 5000, 10000\}$, the parameters are set to $r_{min} = 4$, $\omega_{size} = 29$, $K = 1500$, and $C = 500$. In Section 3.2.4, we will discuss the effects of this parameter selection to the method's performance in detail.

In addition to these parameters, one should select the architecture of the deep belief network. In this work, we fix this architecture. In general, the number of hidden layers determines the abstraction levels represented in the network. We set this number to four. We then select the number of hidden units as 2000, 1000, 500, and 100 from bottom to top layers, having the following considerations. For our work, the hidden unit number in the first layer should be selected large enough to effectively represent the pixels in a local subregion. On the other hand, the number in the last layer should be selected small enough to effectively quantize the subregions. The hidden unit numbers in between should be selected consistent to the selected hidden unit numbers in the first and last layers. The investigation of using different network architectures is considered as future work.

3.2.3 Results

Tables 3.2 and 3.3 report the test set accuracies obtained by our proposed *Deep-Feature* method for the first and second datasets, respectively. These tables provide the class-based accuracies in their first three/five columns and the average

Table 3.2: Test set accuracies of the proposed *DeepFeature* method and the comparison algorithms for the first dataset.

	Norm.	Low	High	Arith. mean	Harm. mean
DeepFeature	98.37	91.59	**98.44**	**96.13**	**96.02**
Handcrafted features					
CooccurrenceMatrix	87.58	84.12	85.60	85.77	85.74
GaborFilter	91.24	82.23	78.60	84.02	83.70
LocalObjectPattern [44]	95.32	92.54	90.27	92.71	92.66
TwoTier [97]	99.18	93.83	93.77	95.59	95.53
Deep learning for supervised classification					
AlexNet	99.39	**97.39**	75.88	90.89	89.53
GoogLeNet	**99.59**	97.04	80.16	92.26	91.40
Inception-v3	**99.59**	93.01	89.11	93.90	93.71
Deep learning for feature extraction (salient points)					
SalientStackedAE	97.35	90.17	93.00	93.50	93.41
SalientConvolutionalAE	96.54	93.96	76.26	88.92	87.94
Deep learning for feature extraction (random points)					
RandomRBM	95.93	87.91	96.89	93.58	93.40
RandomStackedAE [23]	97.96	90.05	90.27	92.76	92.62
RandomConvolutionalAE	95.32	88.63	79.38	87.77	87.28

class-based accuracies in the last two. These tables report the average class-based accuracies instead of the overall test set accuracy since especially the first dataset has an unbalanced class distribution. Here we provide the arithmetic mean of the class-based accuracies as well as their harmonic mean since the arithmetic mean can sometimes be misleading when values to be averaged differ greatly. These results show that the proposed method leads to high test set accuracies, especially for the first dataset. The accuracy for the sub-low-grade cancerous classes decreases, as expected, since this subcategorization is a difficult task even for human observers. The receiver operating characteristic (ROC) curves of these classifications together with their area under the curve (AUC) metrics are reported in Section 3.2.5.

We also compare our method with four groups of other tissue classification algorithms; the comparison results are also provided in Tables 3.2 and 3.3. The first group includes four methods, namely *CooccurrenceMatrix*, *GaborFilter*, *LocalObjectPattern*, and *TwoTier*, that use handcrafted features for image representation.

Table 3.3: Test set accuracies of the proposed *DeepFeature* method and the comparison algorithms for the second dataset.

	Norm.	Low (grade1)	Low (grade1-2)	Low (grade2)	High	Arith. mean	Harm. mean
DeepFeature	96.63	88.83	67.52	62.90	80.54	**79.28**	**77.24**
Handcrafted features							
CooccurrenceMatrix	87.64	71.51	50.43	39.52	78.38	65.50	60.03
GaborFilter	85.96	70.95	22.22	58.06	76.22	62.68	49.47
LocalObjectPattern [44]	92.70	89.39	48.72	58.87	77.30	73.40	69.04
TwoTier [97]	**98.88**	80.45	53.85	62.90	79.46	75.11	71.84
Deep learning for supervised classification							
AlexNet	97.19	**96.09**	35.90	52.42	**87.03**	73.73	63.20
GoogLeNet	97.75	81.56	**76.92**	63.71	61.62	76.31	74.17
Inception-v3	**98.88**	89.94	38.46	**66.94**	86.49	76.14	67.81
Deep learning for feature extraction (salient points)							
SalientStackedAE	98.31	87.71	55.56	58.87	83.24	76.74	72.92
SalientConvolutionalAE	**98.88**	80.45	45.30	51.61	70.27	69.30	63.92
Deep learning for feature extraction (random points)							
RandomRBM	87.08	82.12	56.41	58.87	82.16	73.33	70.88
RandomStackedAE [23]	97.19	82.12	47.01	57.26	82.70	73.26	68.22
RandomConvolutionalAE	96.07	72.63	45.30	44.35	59.46	63.56	58.40

We use them in our comparisons to investigate the effects of learning features directly on image data instead of manual feature definition. The *CooccurrenceMatrix* and *GaborFilter* methods employ pixel-level textures. The *CooccurrenceMatrix* method first calculates a gray-level co-occurrence matrix and then extracts Haralick descriptors from this matrix. The *GaborFilter* method first convolves an image with log-Gabor filters in six orientations and four scales. Then, for each scale, it calculates average, standard deviation, minimum-to-maximum ratio, and mode descriptors on the response map averaged over those of all orientations [34]. Both methods use an SVM with a linear kernel for the final image classification. For both datasets, the proposed *DeepFeature* method leads to test set accuracies much better than these two methods, which employ pixel-level handcrafted features.

The *LocalObjectPattern* [44] and *TwoTier* [97] methods, which we previously developed in our research group, use component-level handcrafted features. The first one defines a descriptor with the purpose of encoding spatial arrangements of the components within the specified local neighborhoods. It is similar to this currently proposed method in the sense that it also represents the components

with circular objects, labels them in an unsupervised way, and uses the labels' distribution for image classification. On the other hand, it uses handcrafted features whereas the currently proposed method uses deep learning to learn the features directly from image data. The comparison results show the effectiveness of the latter approach. The *TwoTier* method decomposes an image into irregular-shaped components, uses Schmid filters [98] to quantify their textures and employs the dominant blob scale metric to quantify their shapes and sizes. At the end, it uses the spatial distribution of these components to classify the image. Although this method gives good results for the first dataset, it is not that successful to further subcategorize low-grade cancerous tissue images (Table 3.3). The proposed *DeepFeature* method also gives the best results for this subcategorization. All these comparisons indicate the benefit of using deep learning for feature extraction.

The second group contains the methods that use CNN classifiers for entire image classification [99, 100, 101, 102]. These methods transfer their CNN architectures (except the last softmax layer since the number of classes is different) and their corresponding weights from the *AlexNet* [103], *GoogLeNet* [104], and *Inception-v3* [104] models, respectively, and fine-tune the model weights on our training images. Since these network models are designed for images with 227×227, 224×224, and 299×299 resolutions, respectively, we first resize our images before using the models. The experimental results given in Tables 3.2 and 3.3 show that the proposed *DeepFeature* method, which relies on characterizing the local salient subregions by deep learning, gives more accurate results than all these CNN classifiers, which are constructed for entire images without considering the saliency.

In the third group of methods (*SalientStackedAE* and *SalientConvolutionalAE*), we extract features from the salient subregions using two other deep learning techniques. Recall that our proposed method trains a deep belief network containing four layers of RBMs and uses the outputs of the RBM in the final layer as the features. We implement these comparison methods to investigate the effectiveness of using an RBM-based feature extractor for this application. The

SalientStackedAE method trains a four-layer stacked autoencoder, whose architecture is the same with our network, and uses the outputs of the final autoencoder as its features. The *SalientConvolutionalAE* method trains a convolutional autoencoder and uses the encoded representation, which is the output of its encoding network, as the features. This convolutional autoencoder network has an encoder with three convolution-pooling layers (with 128, 64, and 32 feature maps, respectively) and a decoder with three deconvolution-upsampling layers (with 32, 64, and 128 feature maps, respectively). Its convolution/deconvolution layers use 3×3 filters and its pooling/upsampling layers use 2×2 filters. Both methods take the RGB values of a subregion as their inputs. Except using a different feature extractor for the salient subregions, the other steps of the methods remain the same. The test set accuracies obtained by these methods are reported in Tables 3.2 and 3.3. When it is compared with *SalientConvolutionalAE*, the proposed *DeepFeature* method leads to more accurate results. The reason might be the following: We use the feature extractor to characterize small local subregions, whose characterizations will later be used to characterize the entire tissue image. The RBM-based feature extractor, each layer of which provides a fully connected network with a global weight matrix, may be sufficient to quantify a small subregion and learning the weights for such a small-sized input may not be that difficult for this application. On the other hand, a standard convolutional autoencoder network, each convolution/deconvolution layer of which uses local and shared connections, may not be that effective for such small local subregions and it may be necessary to customize its layers. The design of customized architectures for this application is considered as future work. The *SalientStackedAE* method, which also uses a fully connected network in each of its layers, improves the results of *SalientConvolutionalAE*, but it still gives lower accuracies compared to our proposed method.

The last group contains three methods that we implement to understand the effectiveness of considering the saliency in learning the deep features. The *RandomRBM* method is a variant of our algorithm. In this method, subregions are randomly cropped out of each image (instead of using the locations of tissue components) and everything else remains the same. Likewise, the *RandomStackedAE*

and *RandomConvolutionalAE* methods are variants of *SalientStackedAE* and *SalientConvolutionalAE*, respectively. They also use randomly selected subregions instead of considering only the salient ones. Note that *RandomStackedAE* uses stacked auto-encoders to define and extract the features, as proposed in [23]. The experimental results are reported in Tables 3.2 and 3.3. The results of all these variants reveal that extracting features from the salient subregions, which are determined by prior knowledge, improves the classification accuracies of their counterparts, especially for the second dataset. All these comparisons indicate the effectiveness of using the proposed RBM-based feature extractor together with the salient points.

The quantitative evaluations provided in Table 3.2 reveal that the *DeepFeature* method leads to higher test set accuracies than all comparisons methods. On the other hand, the test set accuracy of low-grade cancerous class is relatively lower than the test set accuracies of normal and high-grade cancerous classes. In order to improve this accuracy, we decided to classify the test images in the first dataset with the *DeepFeature* method, which is trained with the second dataset, where low-grade cancerous images are recategorized into three separate classes and better represented. Since the second dataset consists of images of five classes, the trained model classifies the test images in the first dataset into five classes. In order to classify the images in the first dataset into three classes, the low-grade (grade1, grade 1-2, and grade2) cancerous classes are annotated with a single low-grade cancerous class. Table 3.4 reports the test set accuracies obtained by the proposed *DeepFeature* method trained on the first and second datasets, respectively. These quantitative results reveal that the proposed *DeepFeature* method trained on the second dataset improves the test set accuracy of the low-grade cancerous class at the expense of decreasing those of the normal and high-grade cancerous classes. The improvement of accuracy in low-grade cancerous images is achieved by training the model with three low-grade cancerous classes.

Table 3.4: Test set accuracies for the first dataset provided by the proposed *DeepFeature* method trained on the first and the second datasets, respectively.

	Norm.	Low	High	Arith. mean	Harm. mean
DeepFeature (trained on the first dataset)	**98.37**	91.59	**98.44**	**96.13**	**96.02**
DeepFeature (trained on the second dataset)	95.93	**96.92**	89.88	94.24	94.14

3.2.4 Parameter Analysis

The *DeepFeature* method has four external parameters: minimum circle radius r_{min}, size of a salient subregion ω_{size}, cluster number K, and SVM parameter C. This section analyzes the effects of the parameter selection on the method's performance. To this end, for each parameter, it fixes the values of the other three parameters and measures the test set accuracies as a function of the parameter of interest. These analyses are depicted in Figures 3.4 and 3.5 for the first and the second datasets, respectively.

The minimum circle radius r_{min} determines the size of the smallest circular object (tissue component) located by the CircleDecomposition algorithm. Its larger values cause not to locate smaller objects, which may correspond to important small tissue components such as nuclei, and not to define salient subregions around them. This may cause an inadequate representation of the tissue, which decreases the accuracy as shown in Figures 3.4(a) and 3.5(a). On the other hand, using smaller values leads to defining noisy objects and the use of the salient subregions around them slightly decreases the accuracy.

The parameter ω_{size} is the size of a salient subregion cropped for each component by the CropWindow algorithm. This parameter determines the locality of the deep features. When ω_{size} is too small, it is not sufficient to accurately characterize the subregion, and thus, the component it corresponds to. This significantly decreases the accuracy. After a certain point, it does not affect the accuracy too much, but of course, increases the complexity of the required deep neural network. This analysis is depicted in Figures 3.4(b) and 3.5(b).

The cluster number K determines the number of labels used for quantizing

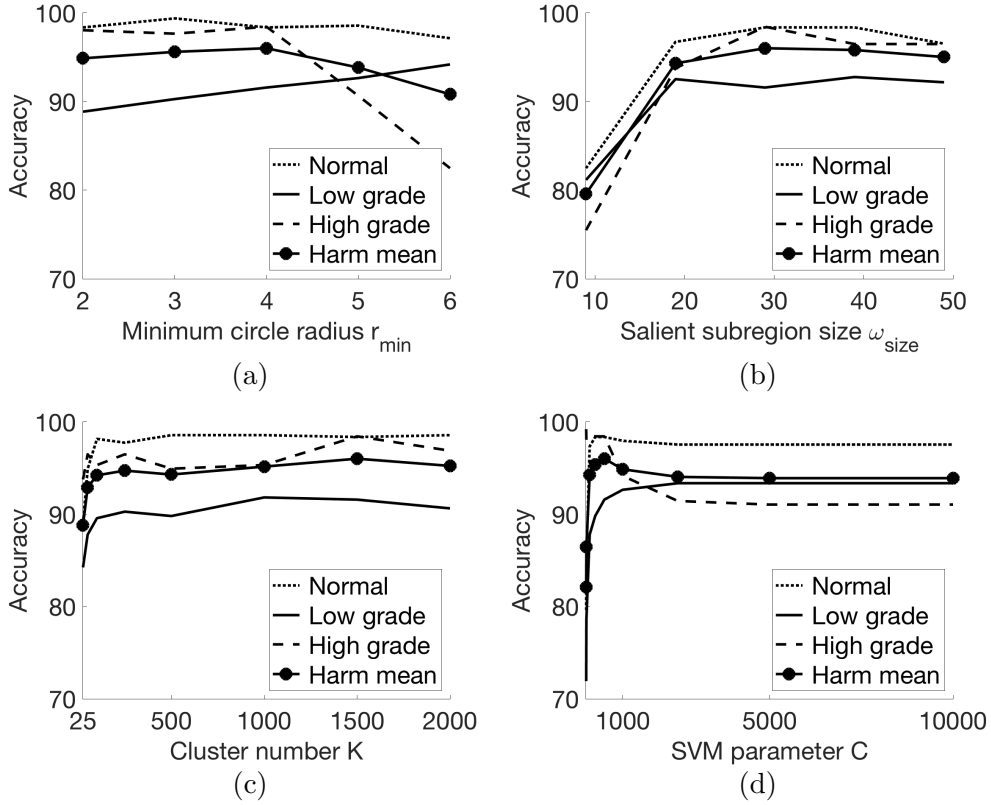


Figure 3.4: For the first dataset, test set accuracies as a function of the model parameters: (a) minimum circle radius r_{min}, (b) size of a salient subregion ω_{size}, (c) cluster number K, and (d) SVM parameter C.

the salient subregions (components). Its smaller values may result in defining the same label for components of different types. This may lead to an ineffective representation, decreasing the accuracy. Using larger values only slightly affects the performance (Figures 3.4(c) and 3.5(c)).

The SVM parameter C controls the trade-off between the training error and the margin width of the SVM model. Using values smaller and larger than necessary may cause underfitting and overfitting, respectively. Unfortunately, similar to many hyperparameters in machine learning, there is no foolproof method for its selection and its value must be determined empirically. As shown in Figures 3.4(d) and 3.5(d), our application necessitates the use of C in the range between 250 and 1000.

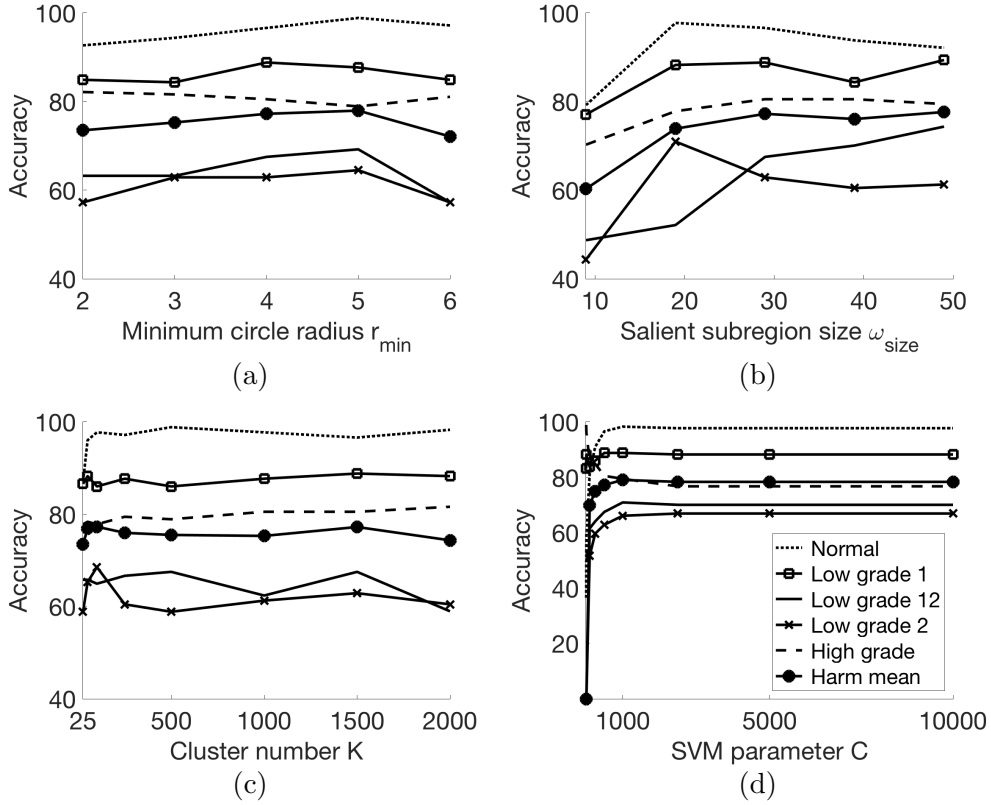


Figure 3.5: For the second dataset, test set accuracies as a function of the model parameters: (a) minimum circle radius r_{min}, (b) size of a salient subregion ω_{size}, (c) cluster number K, and (d) SVM parameter C.

3.2.5 ROC Curves and AUC Analysis

This section presents the ROC curve and AUC analysis for the experiments of our proposed method and the comparison algorithms. Although this analysis is well defined for binary classifications, there is no consensus on how to obtain the ROC curves for multi-class classification problems. In our experiments, we follow the following procedure for both our proposed method and the comparison algorithms. In this procedure, we generate a ROC curve for each class separately, by considering only the posterior probabilities that the multi-class SVM classifier outputs for this particular class (we do not consider the posteriors of the other classes). We threshold these posteriors with the threshold values across the $[0, 1]$ interval and obtain the true positive rate (TPR) and the false positive rate (FPR) for each threshold. We then use these rates to generate the ROC curve.

Table 3.5: For the first dataset, the area under curve (AUC) metrics of the proposed *DeepFeature* method and the comparison algorithms. These metrics are calculated on the test samples of this dataset.

	Norm.	Low	High	Arith. mean	Harm. mean
DeepFeature	0.9974	0.9895	0.9942	0.9937	0.9937
Handcrafted features					
CooccurrenceMatrix	0.9618	0.9615	0.9418	0.9550	0.9549
GaborFilter	0.9728	0.9584	0.9452	0.9588	0.9587
LocalObjectPattern [44]	0.9901	0.9756	0.9841	0.9833	0.9832
TwoTier [97]	0.9996	0.9907	0.9872	0.9925	0.9925
Deep learning for supervised classification					
AlexNet	0.9990	0.9848	0.9750	0.9863	0.9862
GoogLeNet	1.0000	0.9913	0.9859	0.9924	0.9923
Inception-v3	1.0000	0.9882	0.9827	0.9903	0.9902
Deep learning for feature extraction (salient points)					
SalientStackedAE	0.9982	0.9885	0.9888	0.9918	0.9918
SalientConvolutionalAE	0.9984	0.9651	0.9293	0.9643	0.9635
Deep learning for feature extraction (random points)					
RandomRBM	0.9950	0.9837	0.9935	0.9907	0.9907
RandomStackedAE [23]	0.9976	0.9836	0.9811	0.9874	0.9874
RandomConvolutionalAE	0.9927	0.9528	0.9224	0.9560	0.9551

After obtaining the ROC curve for each class separately, we calculate the area under this curve. Tables 3.5 and 3.6 report the class-specific AUC metrics obtained on the test samples of the first and second datasets, respectively. Note that the last two columns of these tables present the averages of these class-specific AUC metrics. Here we provide the arithmetic mean of the class-specific AUC metrics as well as their harmonic mean since the arithmetic mean can sometimes be misleading when values to be averaged differ greatly. These tables indicate the effectiveness of our proposed *DeepFeature* method for the representation and classification of histopathological images. It yields better results than the other algorithms, which is also consistent with our findings reported in Tables 3.2 and 3.3. The ROC curves used in the calculation of these AUC values are presented in Figures 3.6, 3.7 and 3.8 for the first dataset, and in Figures 3.9, 3.10 and 3.11 for the second one.

Table 3.6: For the second dataset, the area under curve (AUC) metrics of the proposed *DeepFeature* method and the comparison algorithms. These metrics are calculated on the test samples of this dataset.

	Norm.	Low (grade1)	Low (grade1-2)	Low (grade2)	High	Arith. mean	Harm. mean
DeepFeature	0.9991	0.9752	0.9284	0.9206	0.9727	0.9592	0.9582
Handcrafted features							
CooccurrenceMatrix	0.9808	0.9083	0.8228	0.7971	0.9541	0.8926	0.8867
GaborFilter	0.9692	0.9100	0.8056	0.8234	0.9483	0.8913	0.8864
LocalObjectPattern [44]	0.9899	0.9622	0.9084	0.8946	0.9612	0.9433	0.9419
TwoTier [97]	0.9997	0.9651	0.8865	0.9001	0.9725	0.9448	0.9427
Deep learning for supervised classification							
AlexNet	0.9974	0.9802	0.8939	0.9132	0.9766	0.9523	0.9505
GoogLeNet	1.0000	0.9893	0.9326	0.8764	0.9764	0.9549	0.9527
Inception-v3	0.9999	0.9773	0.9015	0.9234	0.9677	0.9540	0.9526
Deep learning for feature extraction (salient points)							
SalientStackedAE	0.9998	0.9736	0.9259	0.9130	0.9590	0.9543	0.9532
SalientConvolutionalAE	0.9991	0.9337	0.8539	0.8397	0.9530	0.9159	0.9119
Deep learning for feature extraction (random points)							
RandomRBM	0.9951	0.9588	0.8923	0.9167	0.9693	0.9465	0.9450
RandomStackedAE [23]	0.9993	0.9544	0.8750	0.8894	0.9560	0.9348	0.9325
RandomConvolutionalAE	0.9906	0.9185	0.8549	0.8244	0.9157	0.9008	0.8972

3.2.6 Discussion

This study introduces a new feature extractor for histopathological image representation and presents a system that uses this representation for their classification. This system classifies an image with one of the predefined classes, assuming that it is homogeneous. This section discusses how this system can be used in a digital pathology setup, in which typically lower magnifications are used to scan a slide. Thus, the acquired images usually have a larger field of view and may be homogeneous or heterogeneous. To this end, this section presents a simple algorithm that detects the regions belonging to one of the predefined classes in such a large image. Developing more sophisticated algorithms for the same purpose or for different applications could be considered as future research work.

Our detection algorithm first slides a window with a size that the classification system uses (in our case, the size of 480×640) over the entire large image and then extracts the features of each window and classifies it by the proposed *DeepFeature* method. Since these windows may not be homogeneous, it does not

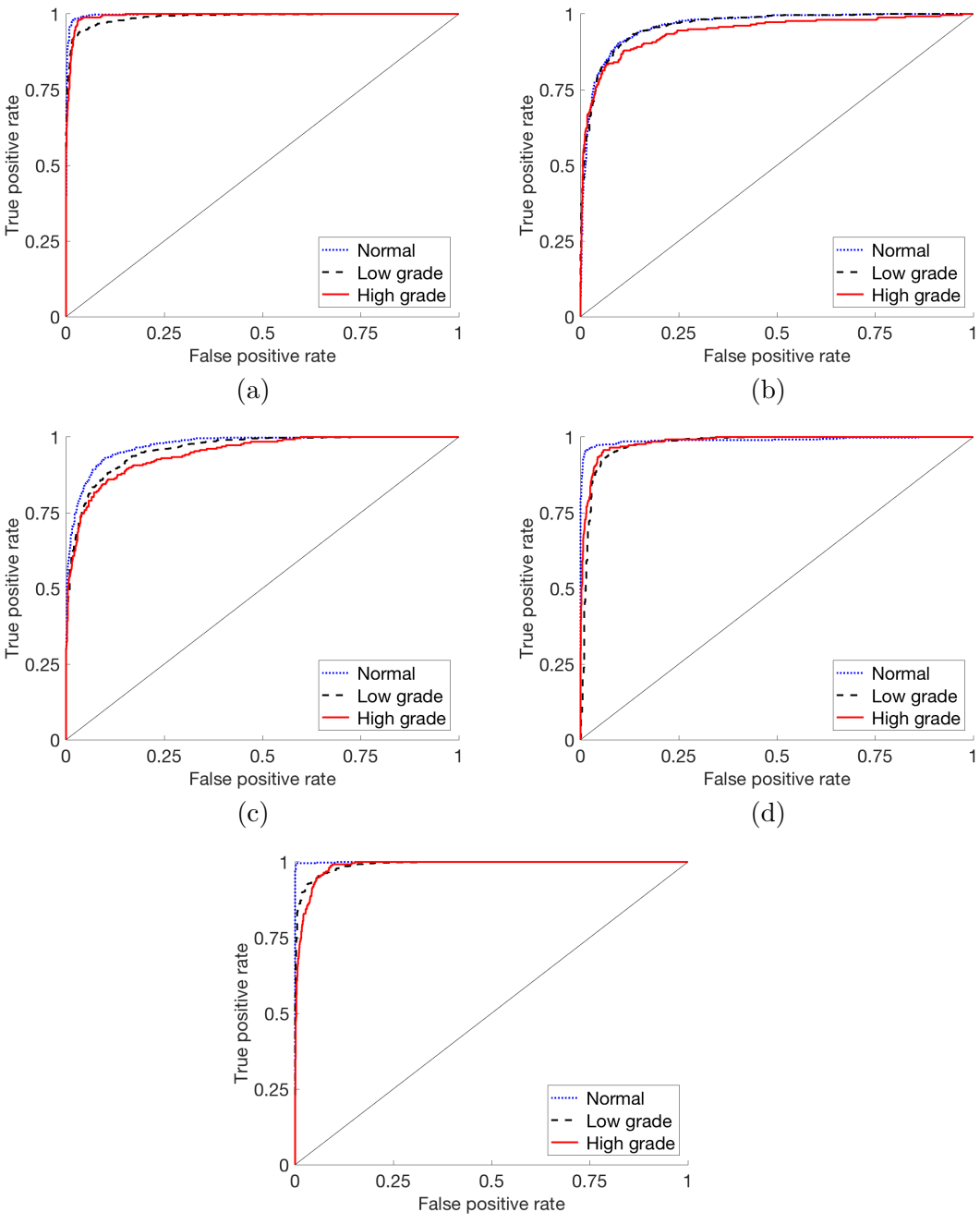


Figure 3.6: ROC curves for the test samples of the first dataset. These curves are obtained for the proposed *DeepFeature* method and the comparison algorithms that use handcrafted features: (a) *DeepFeature*, (b) CooccurrenceMatrix, (c) GaborFilter, (d) LocalObjectPattern [44], and (e) TwoTier [97] methods.

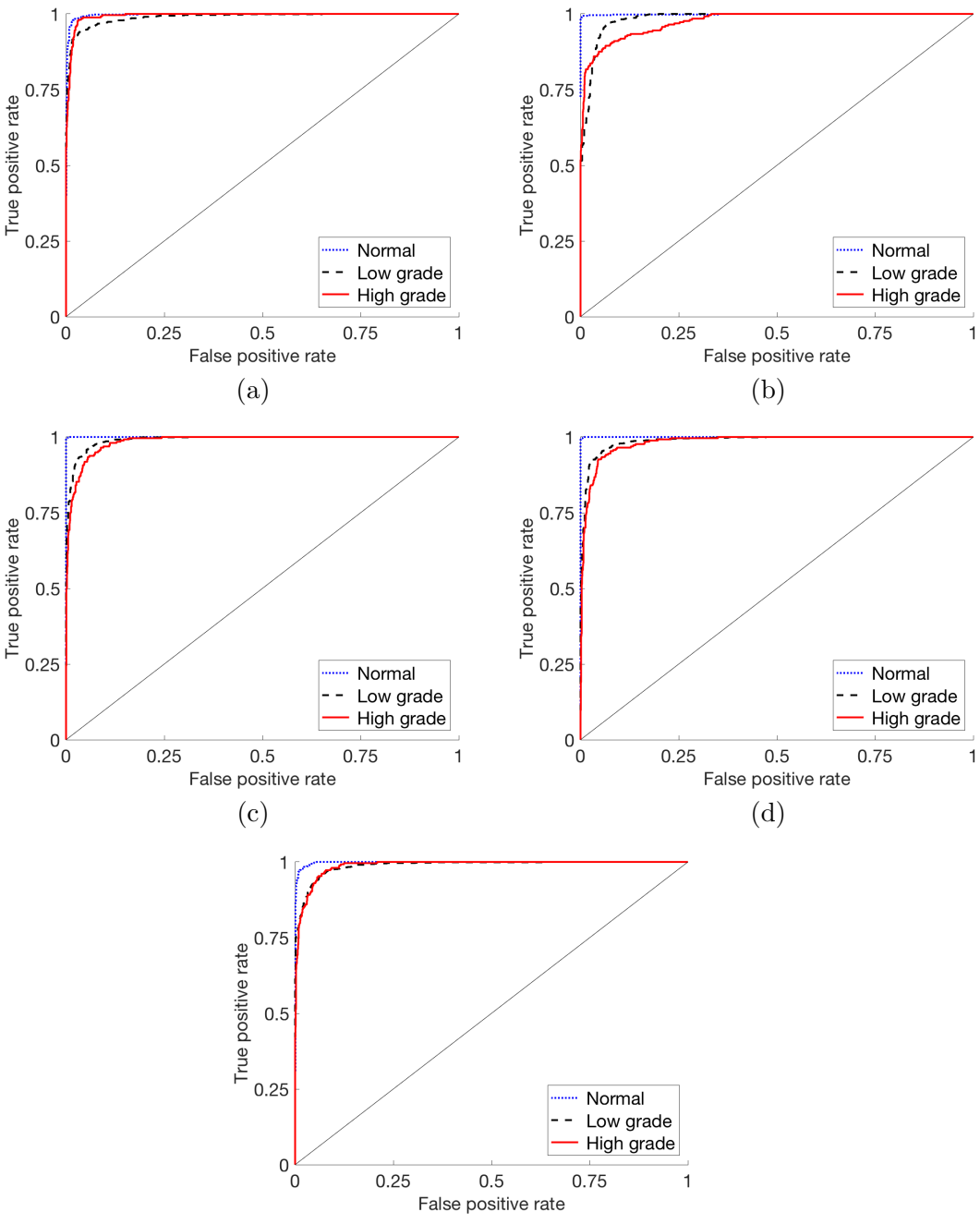


Figure 3.7: ROC curves for the test samples of the first dataset. These curves are obtained for the proposed *DeepFeature* method and the deep learning based comparison algorithms: (a) *DeepFeature*, (b) AlexNet, (c) GoogLeNet, (d) Inception-v3, and (e) SalientStackedAE methods.

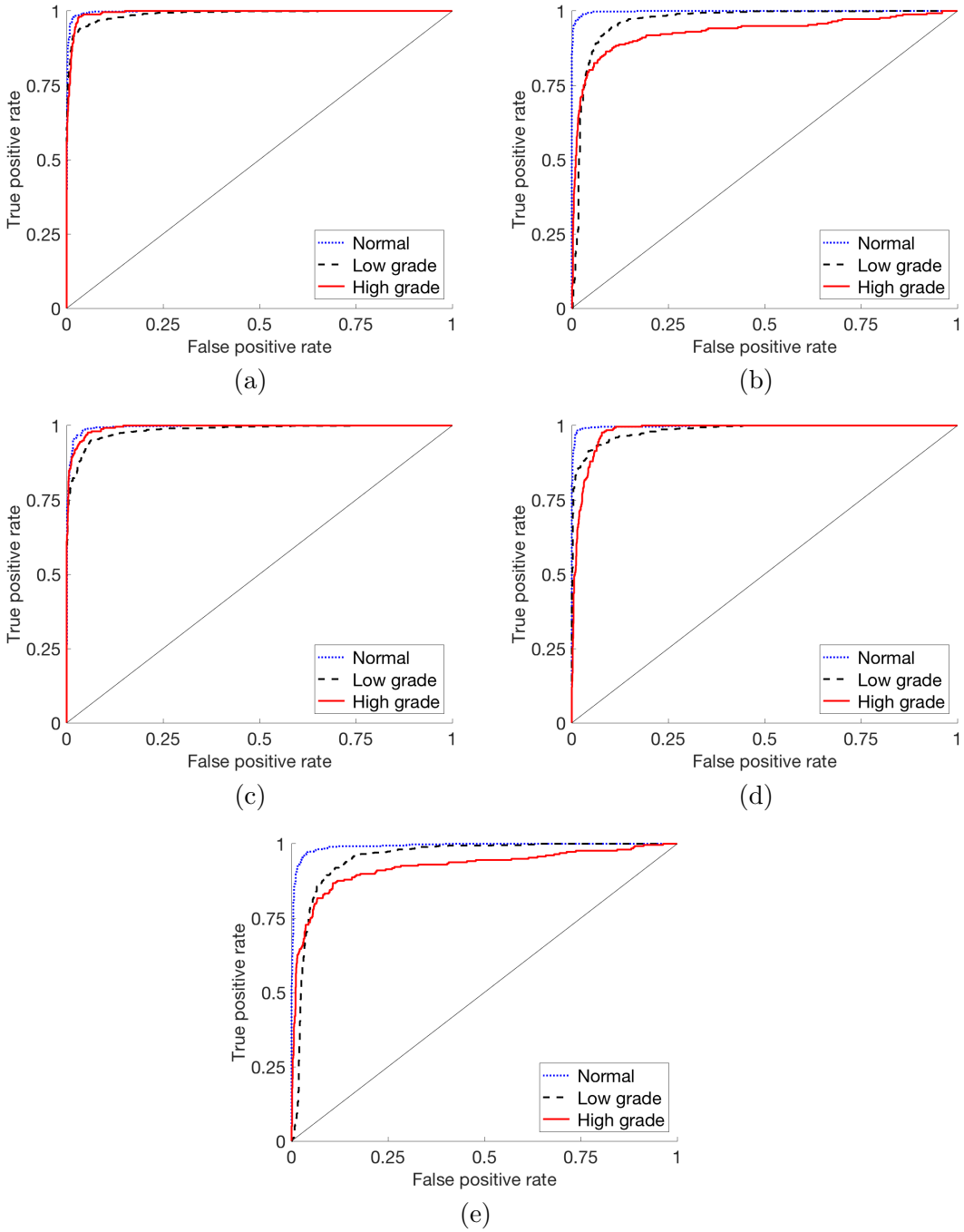


Figure 3.8: ROC curves for the test samples of the first dataset. These curves are obtained for the proposed *DeepFeature* method and the deep learning based comparison algorithms: (a) *DeepFeature*, (b) SalientConvolutionalAE, (c) RandomRBM, (d) RandomStackedAE [23], and (e) RandomConvolutionalAE methods.

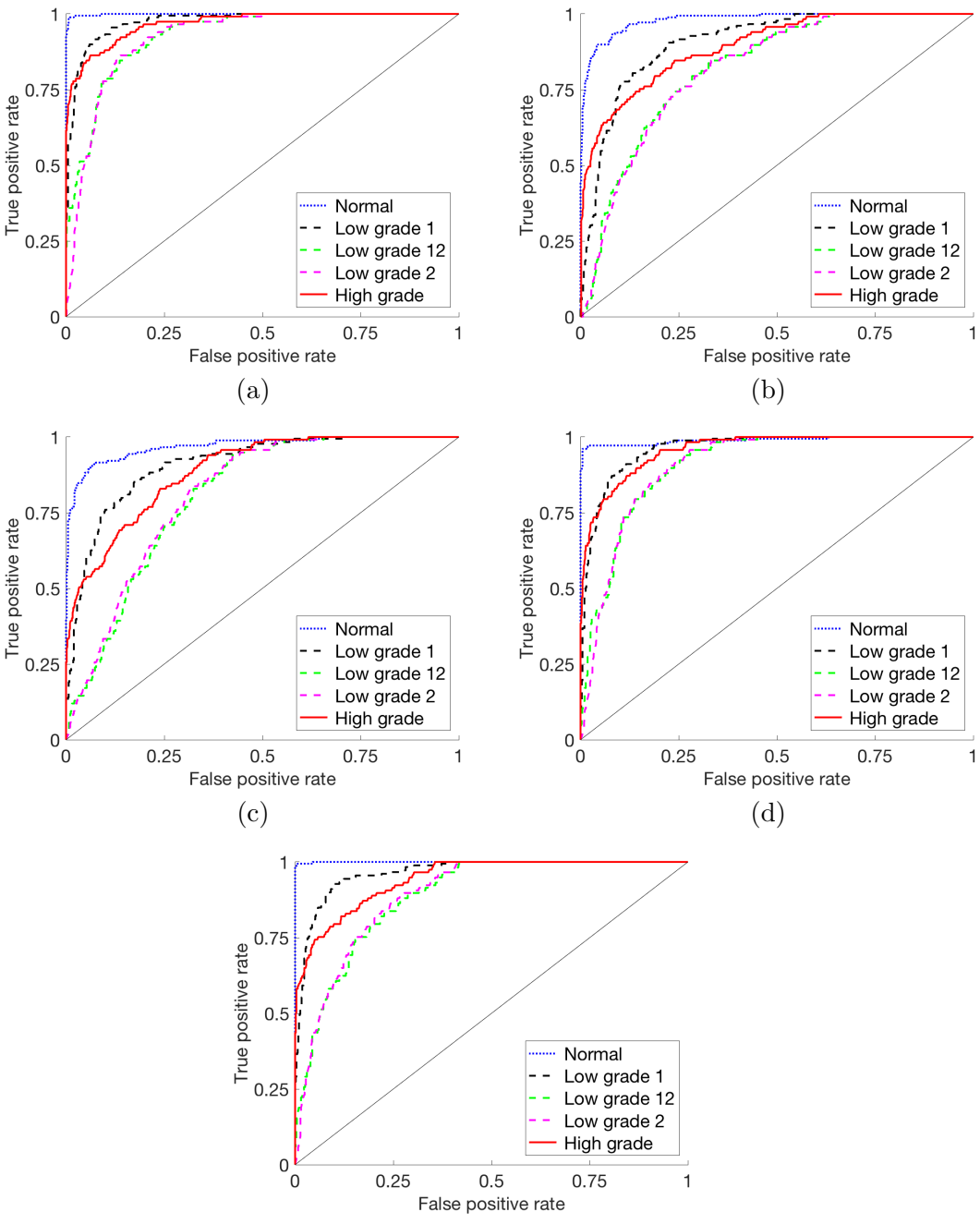


Figure 3.9: ROC curves for the test samples of the second dataset. These curves are obtained for the proposed *DeepFeature* method and the comparison algorithms that use handcrafted features: (a) *DeepFeature*, (b) CooccurrenceMatrix, (c) GaborFilter, (d) LocalObjectPattern [44], and (e) TwoTier [97] methods.

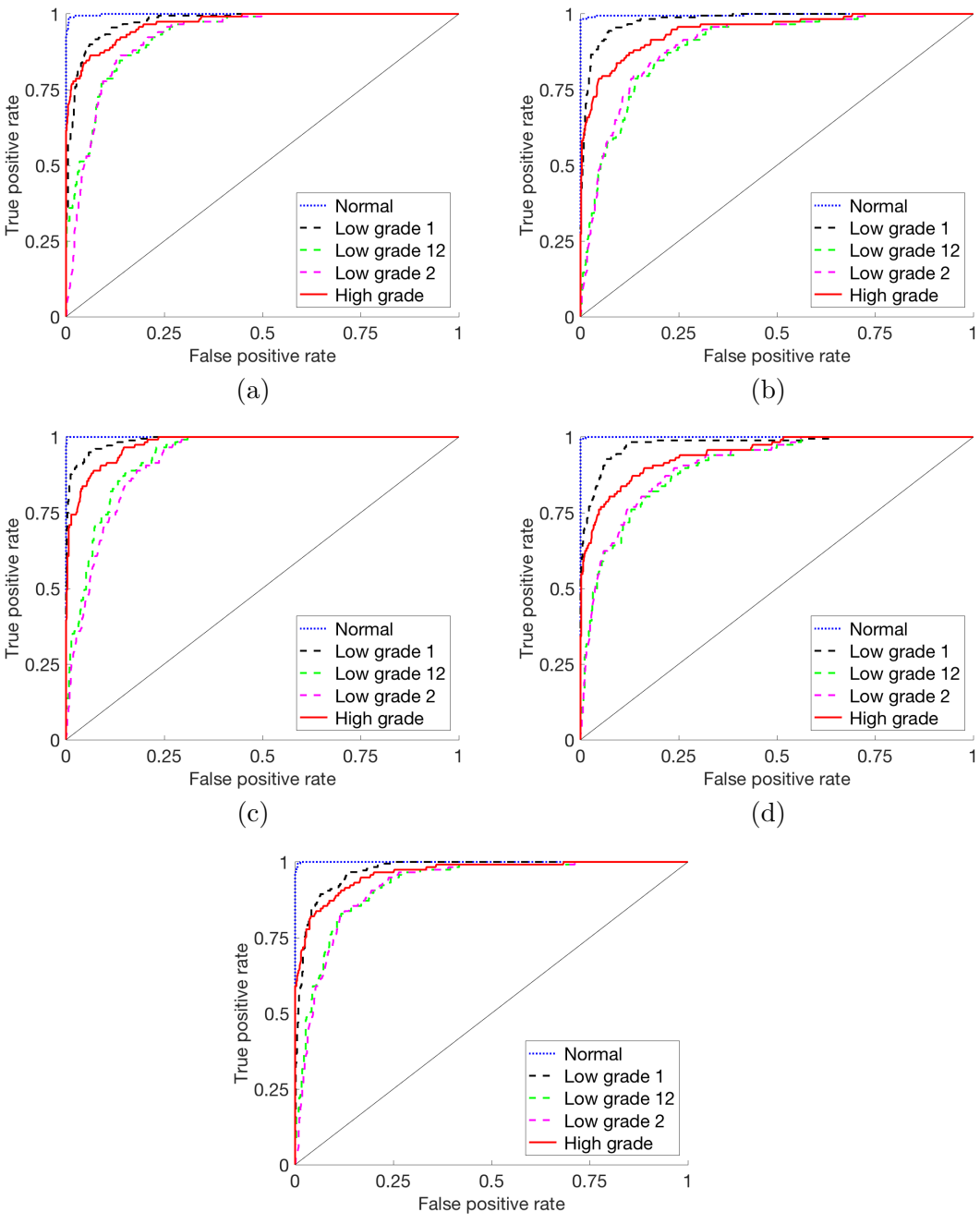


Figure 3.10: ROC curves for the test samples of the second dataset. These curves are obtained for the proposed *DeepFeature* method and the deep learning based comparison algorithms: (a) *DeepFeature*, (b) AlexNet, (c) GoogLeNet, (d) Inception-v3, and (e) SalientStackedAE methods.

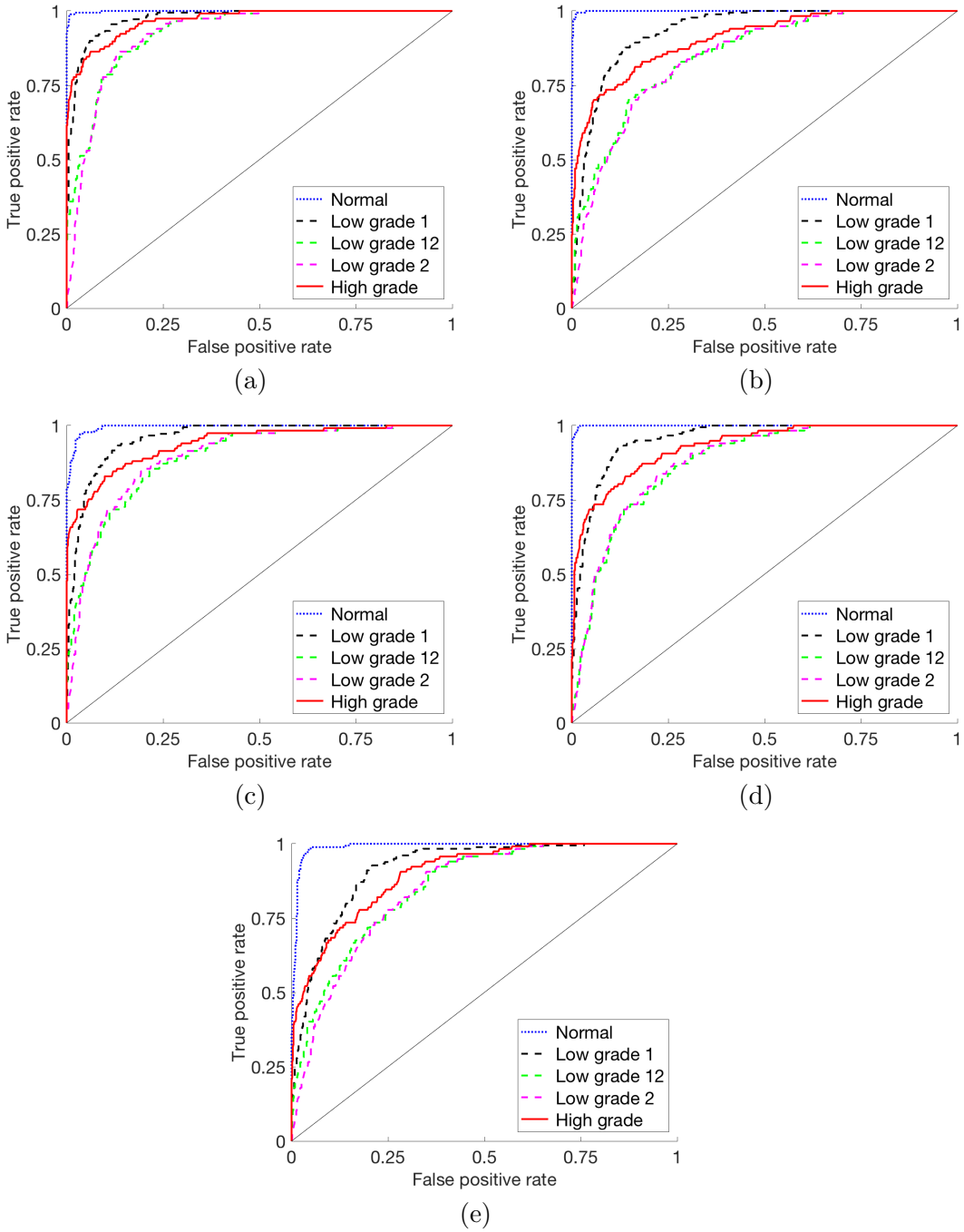


Figure 3.11: ROC curves for the test samples of the second dataset. These curves are obtained for the proposed *DeepFeature* method and the deep learning based comparison algorithms: (a) *DeepFeature*, (b) SalientConvolutionalAE, (c) RandomRBM, (d) RandomStackedAE [23], and (e) RandomConvolutionalAE methods.

directly output the estimated class labels, but instead, it uses the class labels of all windows together with their posteriors in a seed-controlled region growing algorithm. In particular, this detection algorithm has three main steps: posterior estimation, seed identification, and seed growing. All these steps run on circular objects, which we previously define to approximate the tissue components and to represent the salient subregions, instead of image pixels, since the latter is much more computationally expensive. Thus, before starting these steps, the circular objects are located on the large image and the connectivity between them are defined by constructing a Delaunay triangulation on their centroids.

The first step slides a window over the objects and estimates posteriors for all sliding windows by *DeepFeature*. Then, for each object, it accumulates the posteriors of all sliding windows that cover this object. Since our system classifies a window with a predefined class and since these classes may not cover all tissue formations (e.g., lymphoid or connective tissue), this step defines a reject action and assigns it a probability. It uses a very simple probability assignment; the reject probability is 1 if the maximum accumulated posterior is greater than 0.5, and 0 otherwise. The objects are then relabeled by also considering the reject probabilities. As future work, one may define the reject probability as a function of the class posteriors. As an alternative, one may also consider to define classes for additional tissue formations and retrain the classifier. The second step identifies the seeds using the object labels and posteriors. For that, it finds the connected components of the objects that are assigned to the same class with at least T_{seed} probability. It identifies the components containing more than T_{no} objects as the seeds. In our experiments, we set $T_{seed} = 0.90$ and $T_{no} = 500$. The last step grows the seeds on the objects with respect to their posteriors. At the end, the seeds of objects are mapped to image pixels by assigning each pixel the class of its closest seed object, and the seed boundaries are smoothed by majority filtering.

We test this detection algorithm on a preliminary dataset of 30 large images. These images were taken with a $5\times$ objective lens and the image resolution is 1920×2560. Most of the images are heterogeneous; only five of them are homogeneous to test the algorithm also on large homogeneous images. In our tests,

we will directly use the classifier trained for our first dataset without any modification or additional training. Hence, the aim will be to detect low-grade and high-grade colon adenocarcinomatous regions on these large images as well as those containing normal colon glands. Thus, we only annotate those regions on the large images. Example images together with their annotations are given in Figures 3.12, 3.13, and 3.14. The visual results of the algorithm are also given for these examples. For quantitative evaluation, the recall, precision, and F-score metrics are calculated for each class separately. For class C, the standard definitions are as follows: Precision is the percentage of correctly classified C pixels that actually belong to C. Recall is the percentage of actual C pixels that are correctly classified as C by the algorithm. F-score is the harmonic mean of these two metrics. The results for these metrics are reported in Table 3.7. This table also reports the results obtained by relaxing the precision and recall definitions with respect to our application, in which the aim is colon adenocarcinoma detection. Since this cancer type mainly affects epithelial cells, non-epithelial regions are left as unannotated in our datasets. Indeed, one may include these regions to any class without changing the application's aim. Thus, for class C, we relax the definitions as follows: Precision is the percentage of correctly classified C pixels that actually belong to C or a non-epithelial region. Recall is the percentage of actual C pixels that are correctly classified as C or with the reject class by the algorithm.

The visual and quantitative evaluations reveal that the detection algorithm, which uses the proposed classification system, leads to promising results. Thus, it has the potential to be used with a whole slide scanner. To do that, a whole slide should be scanned with a low magnification of the scanner, and the acquired image, which has a larger field of view, can be analyzed by this detection algorithm. Although it yields successful results for many large images, it may also give misclassifications for some of them, especially for those containing relatively large non-epithelial regions; an illustrative example is given in Figure 3.14. When non-epithelial regions are small, incorrect classifications can be compensated by correct classifications of nearby regions and the reject action. However, when

Table 3.7: Results of the colon adenocarcinoma detection algorithm on a preliminary dataset of large images.

	Standard Definitions			**Relaxed Definitions**		
	Precis.	Recall	F-score	Precis.	Recall	F-score
Normal	92.96	79.71	85.83	99.48	88.37	93.60
Low-grade	83.01	91.30	86.96	91.03	93.32	92.16
High-grade	70.82	98.61	82.44	87.00	99.93	93.02

they are large, such compensation may not be possible and the system gives incorrect results since there is no separate class for such regions. Defining an extra class(es) will definitely improve the accuracy on these regions. This is left as future research work.

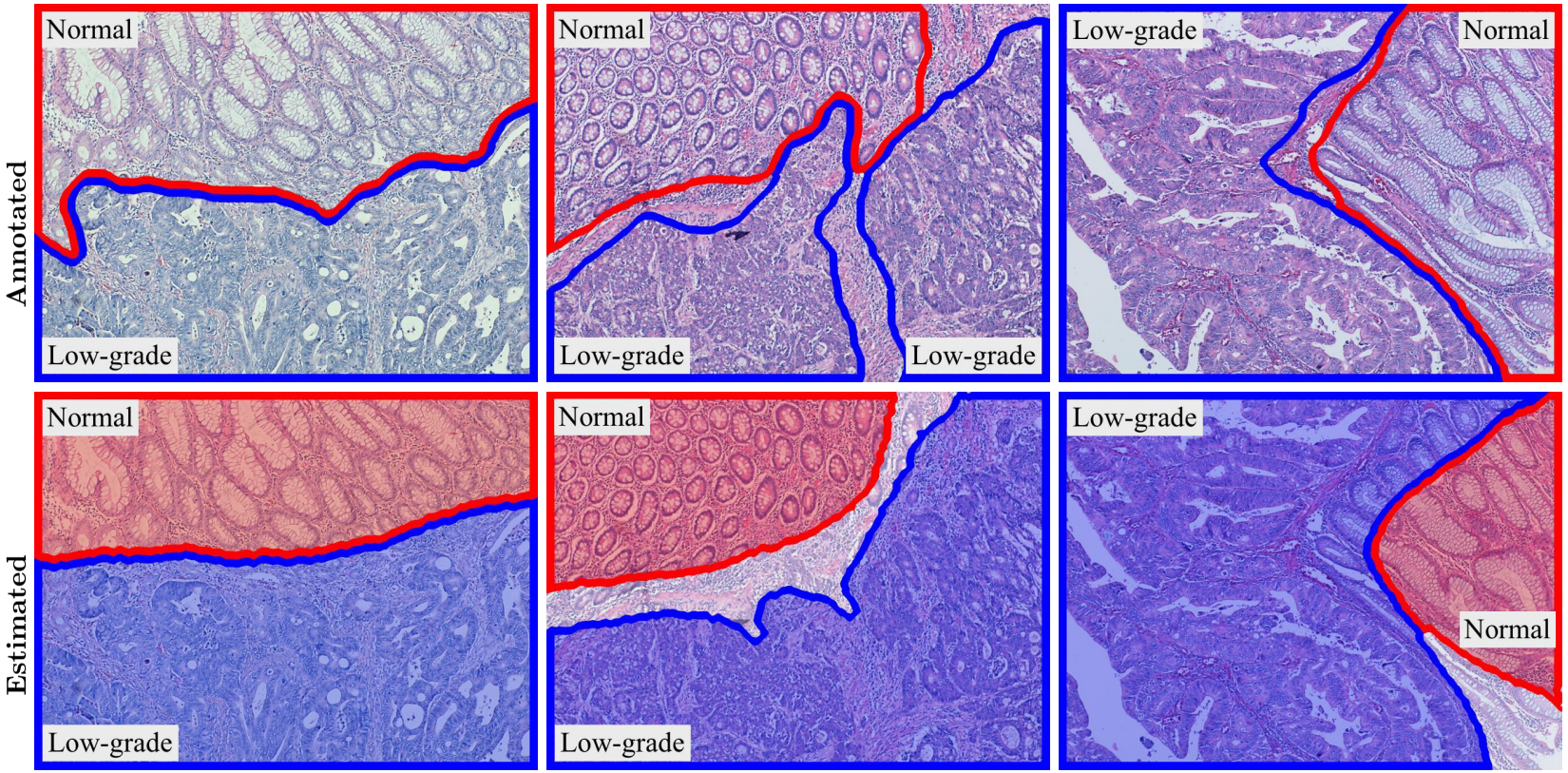


Figure 3.12: Examples of large heterogeneous images together with their visual results obtained by the colon adenocarcinoma detection algorithm. The boundaries of the annotated/estimated normal, low-grade cancerous, and high-grade cancerous regions are shown with red, blue, and green, respectively.

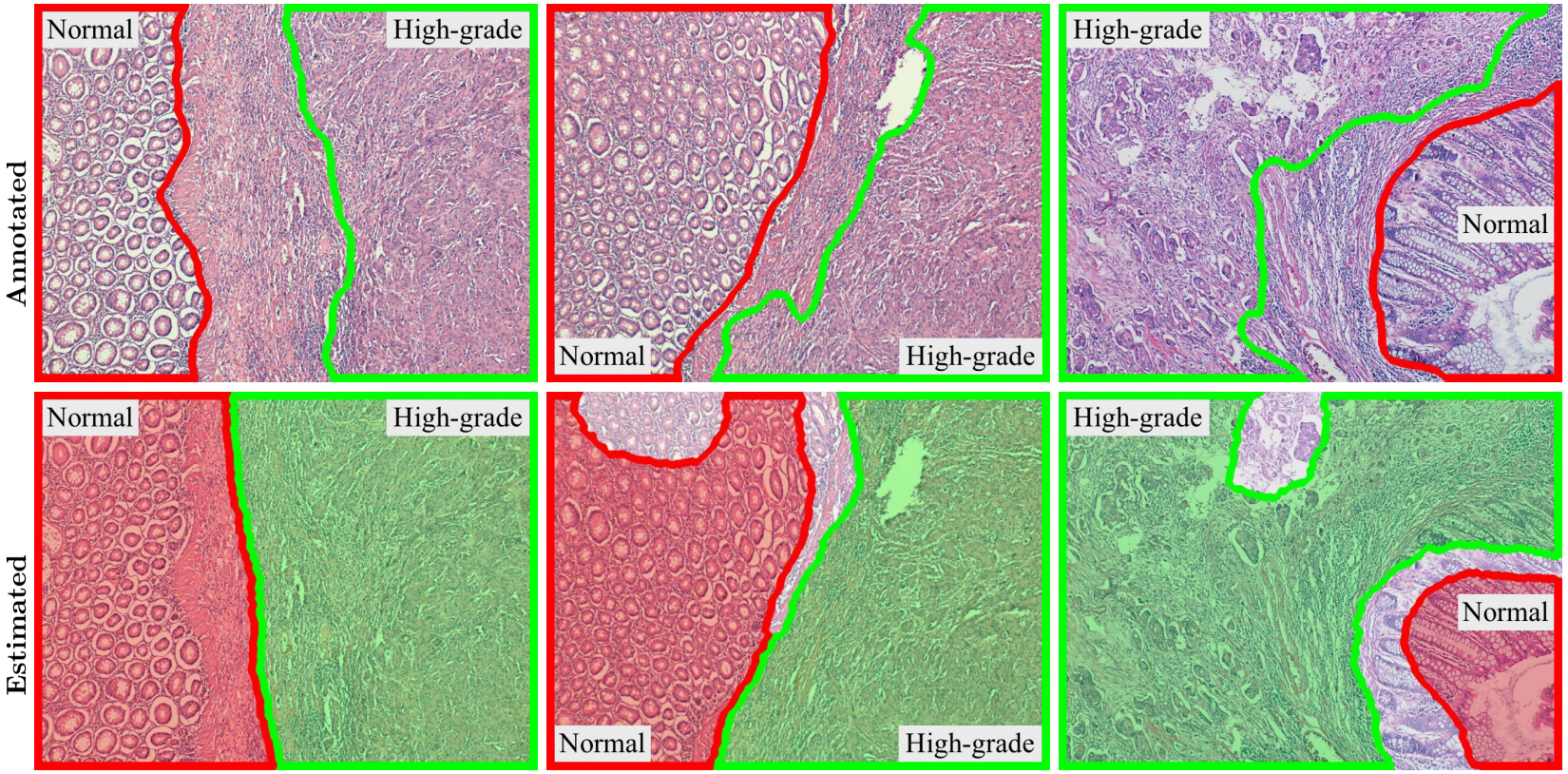


Figure 3.13: Examples of large heterogeneous images together with their visual results obtained by the colon adenocarcinoma detection algorithm. The boundaries of the annotated/estimated normal, low-grade cancerous, and high-grade cancerous regions are shown with red, blue, and green, respectively.

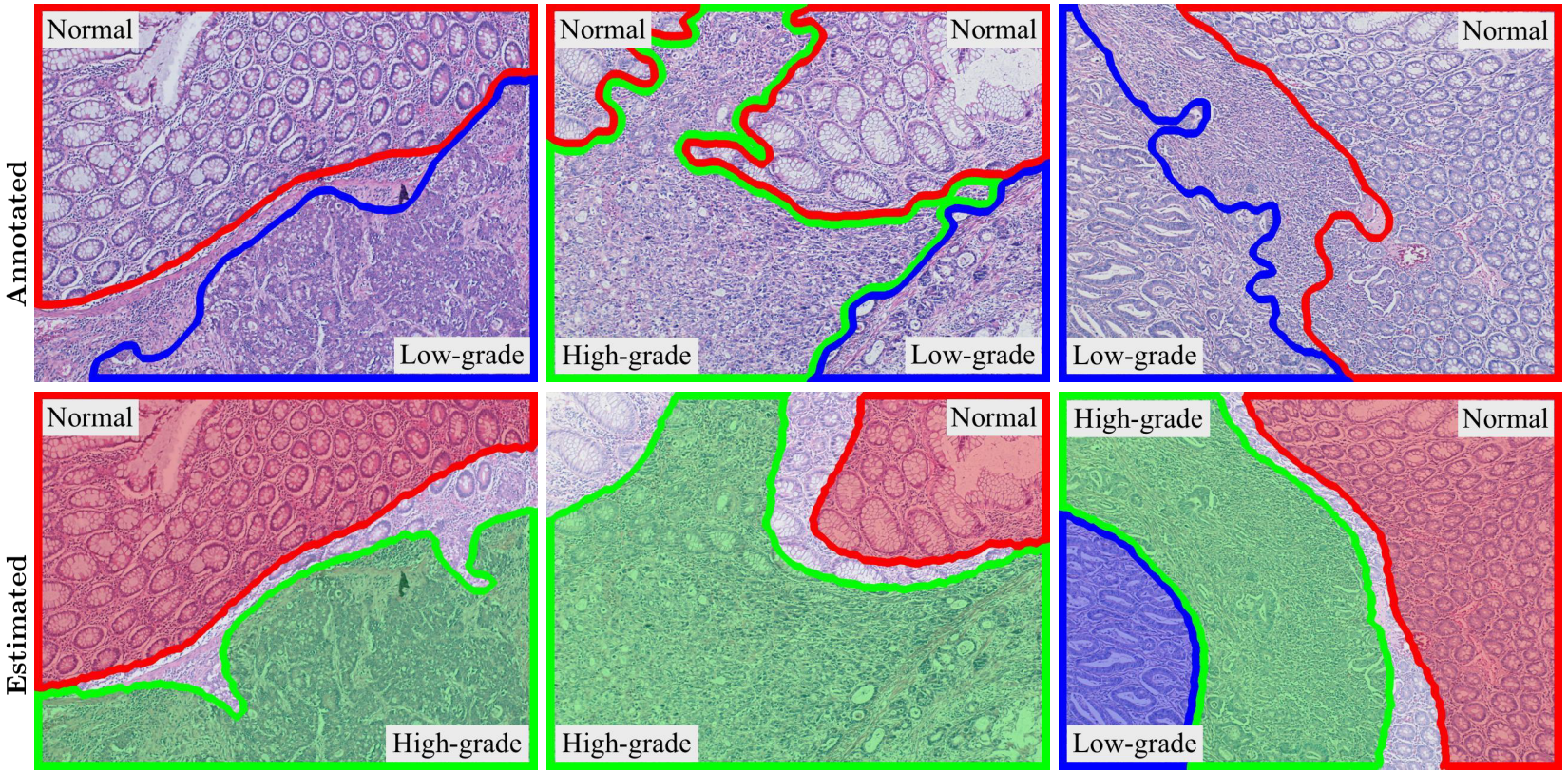


Figure 3.14: Examples of large heterogeneous images that contain regions whose types are incorrectly estimated by our detection algorithm. The boundaries of the annotated/estimated normal, low-grade cancerous, and high-grade cancerous regions are shown with red, blue, and green, respectively.

Chapter 4

Image Embedded Segmentation: Uniting Supervised and Unsupervised Objectives for Segmenting Histopathological Images

4.1 Methodology

The proposed method, which we call *the **iM**age **EM**bedded **S**egmentation (iMEMS) method*, defines a new embedding to transform semantic segmentation to the problem of image-to-image translation and solves it using a conditional generative adversarial network (cGAN). Its motivation is as follows: The proposed transformation facilitates an easy and effective way of uniting a supervised task of semantic segmentation and an unsupervised task of image reconstruction into a single task. By its definition, learning this united task inherently requires meeting the supervised and unsupervised objectives simultaneously. Thus, the network should jointly learn image features to segment an image and context

features to reconstruct it. This joint learning stands as an effective means of regularizing the network training.

The training phase starts with generating a multi-channel output image for each training instance. Then, original input images together with their generated outputs are fed to the cGAN for its training (Figure 4.1). Afterwards, the output of an unsegmented image is estimated by the generator of the trained cGAN. The details are given in the following sections. The iMEMS method is implemented in Python using the Keras framework. The source codes are available at http://www.cs.bilkent.edu.tr/∼gunduz/downloads/iMEMS.

4.1.1 Proposed Embedding

Let I be an RGB image in the training set, G_I be its grayscale, and S_I be its ground truth segmentation map that may contain K possible labels. This embedding generates a K-channel output image O_I by superimposing the grayscale G_I on the segmentation map S_I. For that, for each segmentation label $k \in \{1, ..., K\}$, it generates an output channel $O_I^{[k]}$. For a pixel p, this output channel is defined as follows:

$$O_I^{[k]}(p) = \begin{cases} \left\lfloor \frac{G_I(p)}{2} \right\rfloor + 128 & \text{if } S_I(p) = k \\ 127 - \left\lfloor \frac{G_I(p)}{2} \right\rfloor & \text{if } S_I(p) \neq k \end{cases} \tag{4.1}$$

This definition maps grayscale intensities of all pixels belonging to the k-th label to the interval of [128, 255] in the k-th output channel $O_I^{[k]}$ and to the interval of [0, 127] in all other channels. However, in mapping these intensities to [0, 127], it inverts their values to make the characteristics of pixels in foreground and background regions of the k-th channel more distinguishable. In other words, a grayscale intensity interval [0, 255] is mapped to [128, 255] in the k-th output channel if a pixel belongs to the k-th label, and to [127, 0] otherwise. Note that this definition equally divides the grayscale interval to represent pixels in foreground and background regions in the k-th channel. This is an appropriate choice for our application since each channel needs to represent two types of

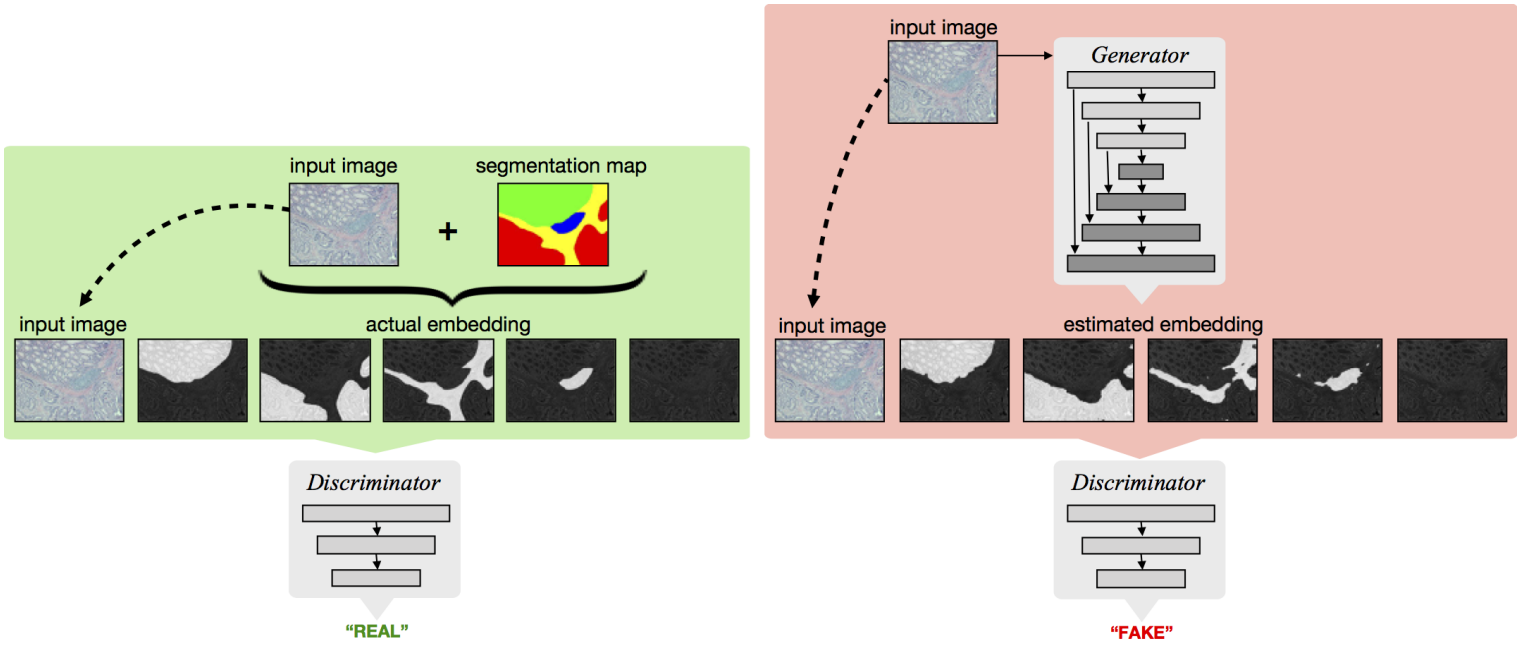


Figure 4.1: Schematic overview of the training phase. It generates a multi-channel output image for each training instance by embedding an input image onto its segmentation map. Each channel corresponds to a segmentation label. Original input images and their generated outputs are fed to the cGAN for its training.

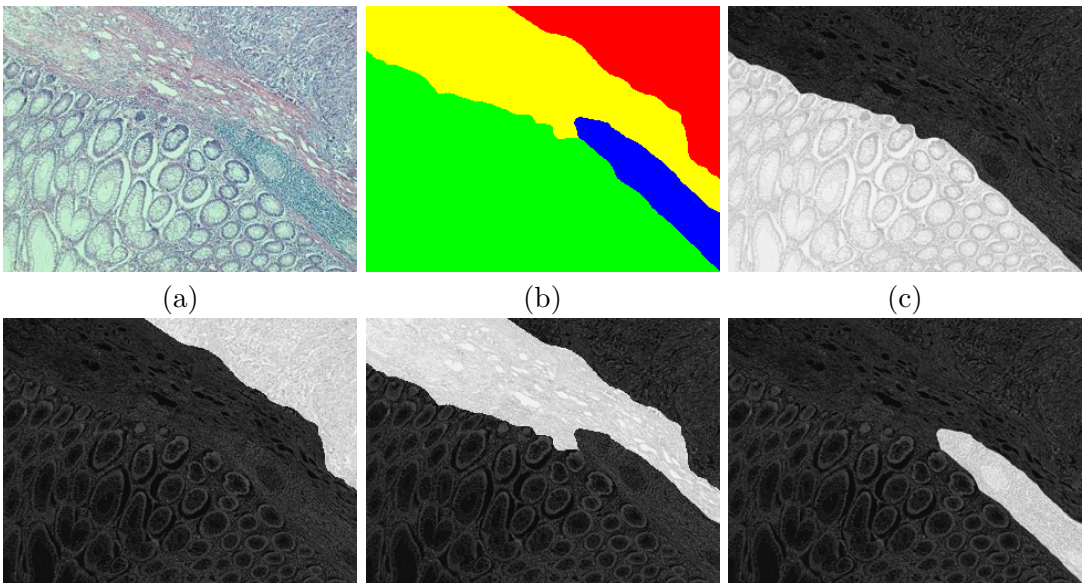


Figure 4.2: (a) An original input image I. (b) Its ground truth segmentation map S_I. (c) The first, (d) second, (e) third, and (f) fourth channels in its output image, which are generated for the segmentation label shown as green, red, yellow, and blue in S_I, respectively. Note that this semantic segmentation problem is a task of predicting one of the five labels for each pixel; this particular image does not contain any pixel belonging to the fifth label. Thus, the generated output image O_I has five channels (i.e., $O_I^{[1]}$, $O_I^{[2]}$, $O_I^{[3]}$, $O_I^{[4]}$, and $O_I^{[5]}$ are generated for the input image). This figure shows only four of these channels.

regions (i.e., background and foreground regions). However, this definition can easily be modified such that it uses unequal divisions of the interval, if this is necessary for other applications.

This definition is illustrated in Figure 4.2. As seen here, foreground regions in each channel seem brighter, as they are mapped to [128, 255], whereas background regions seem darker, as they are mapped to [0, 127]. Thus, it is trivial to segment foreground regions in each channel of this output image. Besides, both foreground and background regions in this output preserve the original image content, which helps regularize a network in learning how to distinguish these two regions.

4.1.2 cGAN Architecture and Training

The definition in Equation 4.1 requires the ground truth map S_I for an input image I. Thus, the iMEMS method only employs this definition to generate the output images for segmented training instances, which are used to train a cGAN. Then, for an unsegmented (test) image, iMEMS estimates this output from an original input image using the trained cGAN. In other words, it translates one image to another using a cGAN.

The generator of this cGAN inputs a normalized RGB image I and outputs a K-channel image $\widehat{O}_I$. It uses a UNet architecture with an encoder and a decoder connected by symmetric connections (Figure 4.3). The convolution layers, except the last one, use 3×3 filters and the ReLU activation function. The last layer uses a linear function since it estimates continuous intensity values of the output image. The pooling/upsampling layers use 2×2 filters. Extra dropout layers are added to reduce overfitting; the dropout factor is set to 0.2.

The discriminator inputs a normalized RGB image and the K-channel output image corresponding to this input. Its output is a class label to indicate whether the output image is real or fake; i.e., it estimates if this output is calculated by Equation 4.1 using the ground truth or produced by the generator. Its architecture is given in Figure 4.4. It has the same operations with the generator's encoder except that its last layer uses the sigmoid function. This network uses a convolutional PatchGAN classifier [89], which uses local patches to determine whether the output image is real or fake rather than the entire image.

The generator and discriminator networks are trained from scratch. The batch size is 1. The network weights are learned on the training images for 300 epochs. At each epoch, the loss is calculated on the validation images and the network that gives the minimum validation loss is selected at the end.

The loss settings of this cGAN are the same with [89]. The objective function is $\text{argmin}_G \text{max}_D \ \mathcal{L}_{adv}(G, D) + \lambda \ \mathcal{L}_{L1}(G)$, where $\mathcal{L}_{adv}(G, D)$ is the adversarial loss on the discriminator's outputs and $\mathcal{L}_{L1}(G)$ is the L1 loss on the generator's

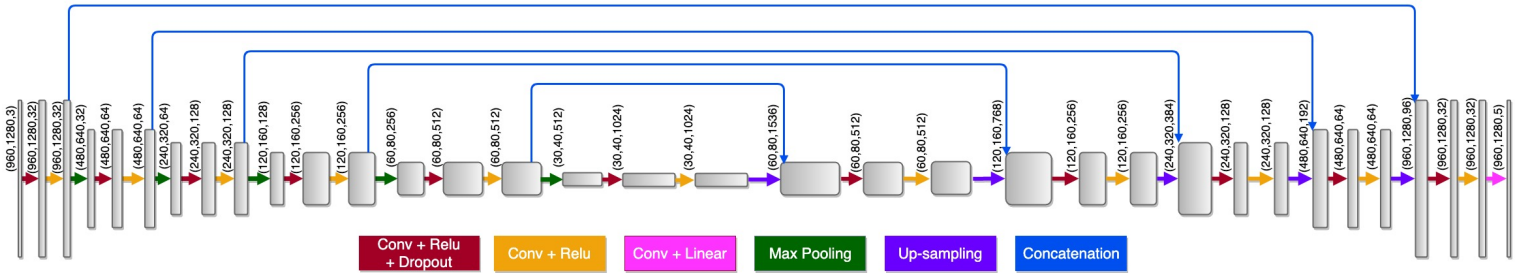


Figure 4.3: Architecture of the generator network in the cGAN. Different layers and operations are indicated with different colors. The resolution of the feature maps in each layer together with the number of these feature maps are also indicated.

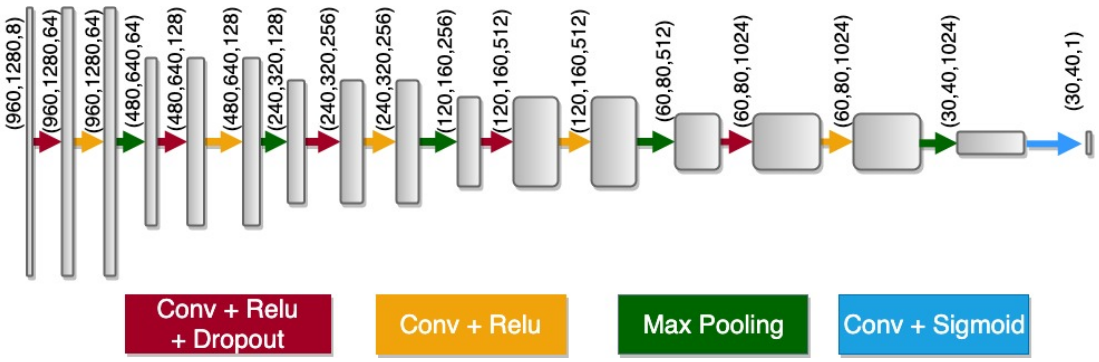


Figure 4.4: Architecture of the discriminator network in the cGAN.

output. Similar to [89], the weight λ of the L1 loss is selected as 100. It is worth to noting that although this objective linearly combines two losses, its purpose is different than the proposed iMEMS method. As opposed to iMEMS, this objective does not directly aim to combine the losses of the supervised task of semantic segmentation and the unsupervised task of image reconstruction. The iMEMS method defines an embedding to unite these two tasks into a single one and uses a cGAN for better learning this united task. Indeed, both the generator and the discriminator of the cGAN define their tasks on the united task of the iMEMS method, which means the adversarial and L1 losses are also defined on this united task.

4.1.3 Tissue Segmentation

For an unsegmented image U, the iMEMS method estimates the output $\widehat{O}_U$ using the generator of the trained cGAN and segments it based on this estimated output. In particular, it classifies each pixel p with a segmentation label k whose corresponding output has the highest estimated value; that is, $\widehat{S}_U(p) = \text{argmax}_k \widehat{O}_U^{[k]}(p)$. For the image shown in Figure 4.2, the estimated output images are illustrated in Figure 4.5.

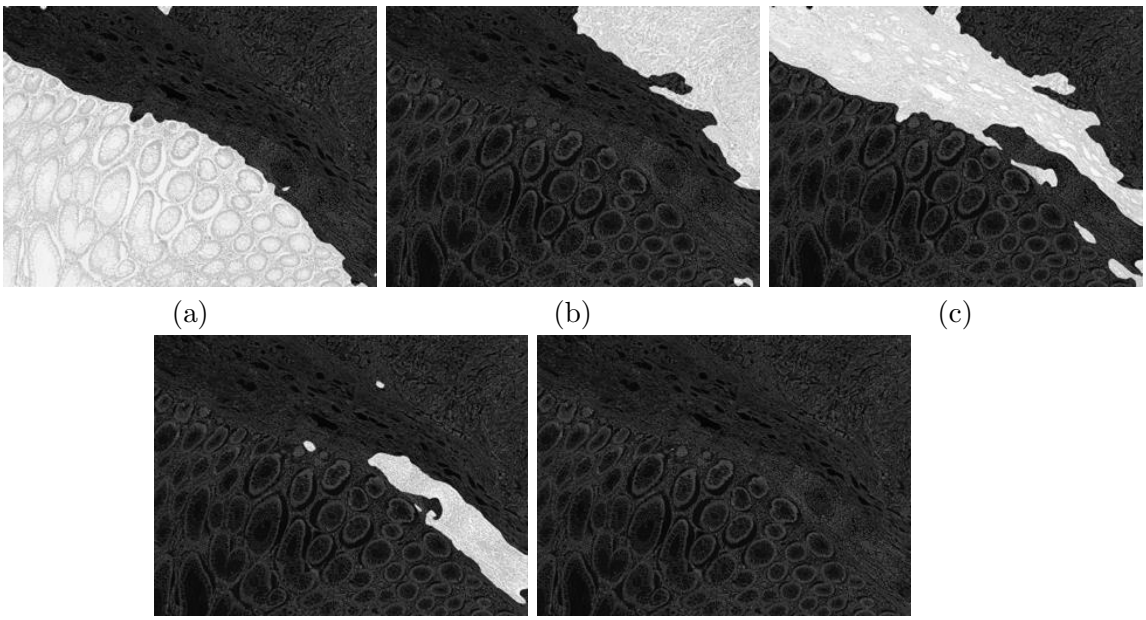


Figure 4.5: Output maps $\widehat{O}_U^{[k]}$ estimated by the generator of the cGAN for the image shown in Figure 4.2.

4.2 Experiments

4.2.1 Datasets

We test the iMEMS method on three datasets that contain microscopic images of hematoxylin-and-eosin stained tissues. The first one is an in-house colon dataset and the other two are publicly available epithelium and tubule datasets, which are prepared by another research group [22].

The ***in-house dataset*** contains 365 images of colon tissues collected from the Pathology Department Archives of Hacettepe University. Images are scanned at $5\times$, using a Nikon Coolscope Digital Microscope. Image resolution is 960×1280. In each image, regions are annotated considering five labels. The details of this annotation are given in Section 4.2.1.1. In this dataset, 100 images are randomly selected as training instances. The remaining ones are used as test instances, on which we measure the performance of our method and comparison algorithms.

The ***epithelium dataset*** consists of 42 estrogen receptor positive breast cancer images scanned at 20×. Image resolution is 1000 × 1000. In each image, non-overlapping regions are annotated as either epithelium or background [22]. Since the size of this dataset is relatively small, we randomly split it into five folds and measure the performance using five-fold cross-validation. Furthermore, we divide an image of each fold into four equal non-overlapping parts in order to make images optimal for the proposed architecture and also to increase the number of training instances. Note that all four parts belonging to the same image are used in the same fold.

The ***tubule dataset*** consists of 85 colorectal images scanned at 40×. As these images have different resolutions, we rescale them to 522 × 775 pixels, which is the resolution of more than 90 percent of all images. In each image, tubule and background regions are annotated [22]. Likewise, the size of this dataset is also relatively small. Thus, we also use five-fold cross-validation to assess the methods' performance.

4.2.1.1 Annotation Procedure for In-House Colon Dataset

In each image, non-overlapping regions are annotated with one of the five labels: normal, tumorous (colon adenocarcinomatous), connective tissue, dense lymphoid tissue, and non-tissue (empty glass and debris). This annotation is not perfect and may contain inevitable inconsistencies since small subregions of different labels may be found together, due to the nature of colon tissues, and their separate annotation may become quite difficult at the selected magnification. Considering the following three factors that mainly contribute to this difficulty, images are annotated as consistently as possible.

First, normal/tumorous regions consist of small connective tissue and non-tissue subregions. This is inevitable since a normal/tumorous region contains colon glands, which have a luminal area (empty looking subregion) inside, and connective tissue as the supporting material between the glands. In annotations, such luminal areas and connective tissues are included into the corresponding

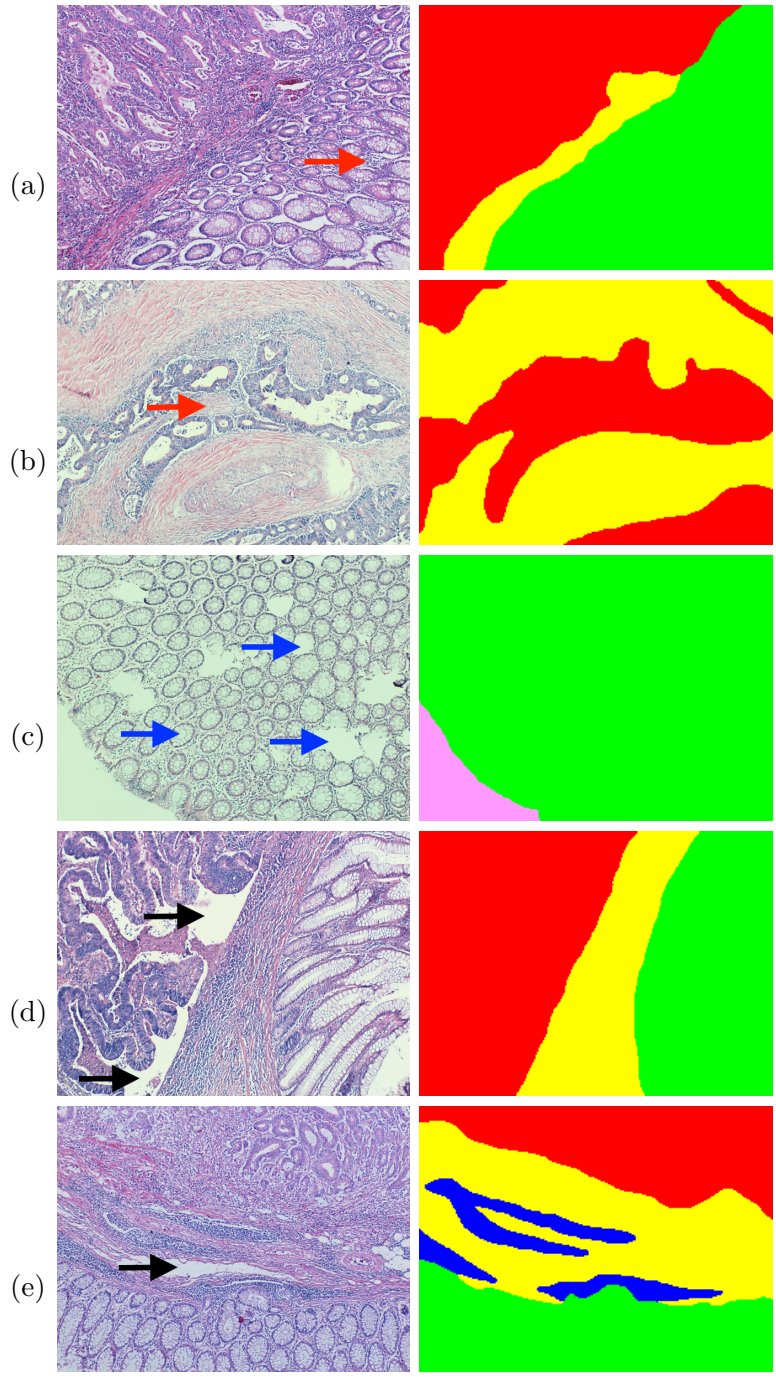


Figure 4.6: Example images of our in-house colon dataset together with their annotations. In annotations, each label is shown with a different color: normal (green), tumorous (red), connective tissue (yellow), dense lymphoid tissue (blue), and non-tissue (pink).

normal/tumorous region. However, if there exists a "wide" enough connective tissue region between the glands, it is separately annotated with the connective tissue label. In Figures 4.6(a) and 4.6(b), two such small connective tissue sub-regions are indicated with red arrows. They are included in their corresponding normal and tumorous regions since they are relatively small. On the other hand, wider connective tissues are annotated as separate regions (yellow regions shown in the second row). Here we make every effort to be as consistent as possible to identify wide regions. Likewise, in Figure 4.6(c), the normal region contains small empty (non-tissue) parts, some of which are shown with blue arrows. These small parts are included into the normal region. However, the left-bottom corner of the image is annotated as a separate region since it belongs to the empty glass but not the tissue.

Second, due to the density heterogeneity in a colon tissue, sectioning paraffin-embedded tissue blocks may result in white artifacts. Examples are shown with black arrows in Figures 4.6(d) and 4.6(e). When these artifacts are found next to a gland, they are included into the normal/cancerous region that the gland belongs to. Otherwise, they are included into the corresponding connective tissue region. Third, lymph cells are found almost everywhere in the tissue. The group of these cells is only annotated as a separate region when they form a dense lymphoid tissue, see Figure 4.6(e). Likewise, we make every effort to be consistent to identify the dense regions.

4.2.2 Results

Two metrics are used for quantitative evaluation. The first one is the pixel-level accuracy, which gives the percentage of correctly predicted pixels in all images. The second one is the pixel-level F-score that is calculated for each segmentation label separately. That is, for each label, the F-score is calculated considering the pixels of this label as positive and those of the other label(s) as negative. The average of these class-wise F-scores is also calculated. The quantitative results are reported in Table 4.1. In this table, the metrics are calculated on the test set

images for the in-house colon dataset. For the other two datasets, these are the average test fold metrics calculated over five runs (using five-fold cross-validation). Note that, for each run, the method of interest is trained on the images of four out of five folds and the remaining one is considered as the test fold. These results show that the proposed iMEMS method gives high F-scores for all segmentation labels, leading to the best accuracy and the best average F-score, for all datasets.

Visual results on example test set/fold images are shown in Figures 4.7-4.17. They reveal that the iMEMS method does not only give higher performance metrics but also produces more realistic segmentations that adhere to spatial contiguity in pixel predictions, especially for the in-house colon dataset (Figures 4.7-4.11). This is attributed to the effectiveness of using the proposed embedding as the output and learning it with a cGAN. Since this output also includes the original image content, it provides regularization on the segmentation task. Moreover, since the discriminator performs real/fake classification on the entire output, it enforces the generator to produce embeddings that better preserve the shapes of the segmented regions.

To better explore these two factors (namely, using the proposed embedding and learning it with a cGAN), we compare iMEMS with five comparison algorithms summarized in Table 4.2. These algorithms either estimate the original segmentation map or the proposed embedding using either a UNet or a cGAN. For fair comparisons, the algorithms that use a cGAN have the same architecture with our method and those that use a UNet have the architecture of our method's generator. The last layer of a network uses a linear function if it estimates the proposed embedding, and a softmax function if it estimates the segmentation map. Two comparison algorithms use a multi-task network that concurrently learns the segmentation and image reconstruction tasks. These networks contain a shared encoder and two parallel decoders, whose architectures are the same with those of the generator.

First, we compare iMEMS with three algorithms that consider none or only one of the two factors. `UNet-C-single` is the baseline that considers none; it

Table 4.1: F-scores and accuracies of the proposed iMEMS method and the comparison algorithms. (a) For the in-house colon dataset, these metrics are obtained on the test set. (b) For the epithelium dataset, these are the average metrics obtained on the five test folds. (c) For the tubule dataset, these are also the average metrics obtained on the five test folds.

	F-scores						**Accuracy**
	Normal	*Tumorous*	*Connective*	*Lymphoid*	*Non-tissue*	*Average*	
iMEMS	**94.81**	**93.12**	**84.43**	**80.54**	**86.00**	**87.78**	**91.76**
`UNet-C-single`	92.89	91.83	79.96	77.55	61.28	80.70	89.27
`cGAN-C-single`	92.45	90.65	76.74	78.87	80.33	83.81	88.49
`UNet-R-single`	93.12	91.17	75.72	72.78	78.29	82.22	88.80
`UNet-C-multi`	92.89	91.85	82.00	78.91	77.83	84.70	89.91
`UNet-C-multi-int`	90.43	89.80	80.49	79.13	83.09	84.59	88.03

(a)

	F-scores			**Accuracy**
	Epithelium	*Backgr.*	*Average*	
iMEMS	**85.51**	**92.40**	**88.96**	**90.17**
`UNet-C-single`	81.86	89.74	85.80	86.96
`cGAN-C-single`	81.67	90.14	85.91	87.26
`UNet-R-single`	82.59	91.02	86.81	88.20
`UNet-C-multi`	81.82	90.65	86.23	87.72
`UNet-C-multi-int`	81.71	90.57	86.14	87.60

(b)

	F-scores			**Accuracy**
	Tubule	*Backgr.*	*Average*	
iMEMS	**87.08**	**87.00**	**87.04**	**87.09**
`UNet-C-single`	84.65	83.57	84.11	84.26
`cGAN-C-single`	85.01	84.58	84.79	84.88
`UNet-R-single`	84.48	83.83	84.15	84.37
`UNet-C-multi`	86.06	84.47	85.27	85.43
`UNet-C-multi-int`	85.67	85.30	85.49	85.59

(c)

Table 4.2: Summary of the algorithms used for the comparative study. The naming convention of these algorithms is X-Y-Z. X is the network type that the algorithm uses. Y is R (regression) if the estimated output is the proposed embedding and C (classification) if it is the segmentation map. Z indicates whether the algorithm uses a single-task or a multi-task network.

Method name	Network	Output	Task
iMEMS	cGAN	Proposed embedding	Single-task regression
`UNet-C-single`	UNet	Segmentation map	Single-task classification
`cGAN-C-single`	cGAN	Segmentation map	Single task classification
`UNet-R-single`	UNet	Proposed embedding	Single-task regression
`UNet-C-multi`	UNet	Segmentation map and reconstructed image	Multi-task classification and image reconstruction (*reconstruction loss is calculated at the input level*)
`UNet-C-multi-int`	UNet	Segmentation map and reconstructed image	Multi-task classification and image reconstruction (*reconstruction loss is calculated at the input level as well as the intermediate layers*)

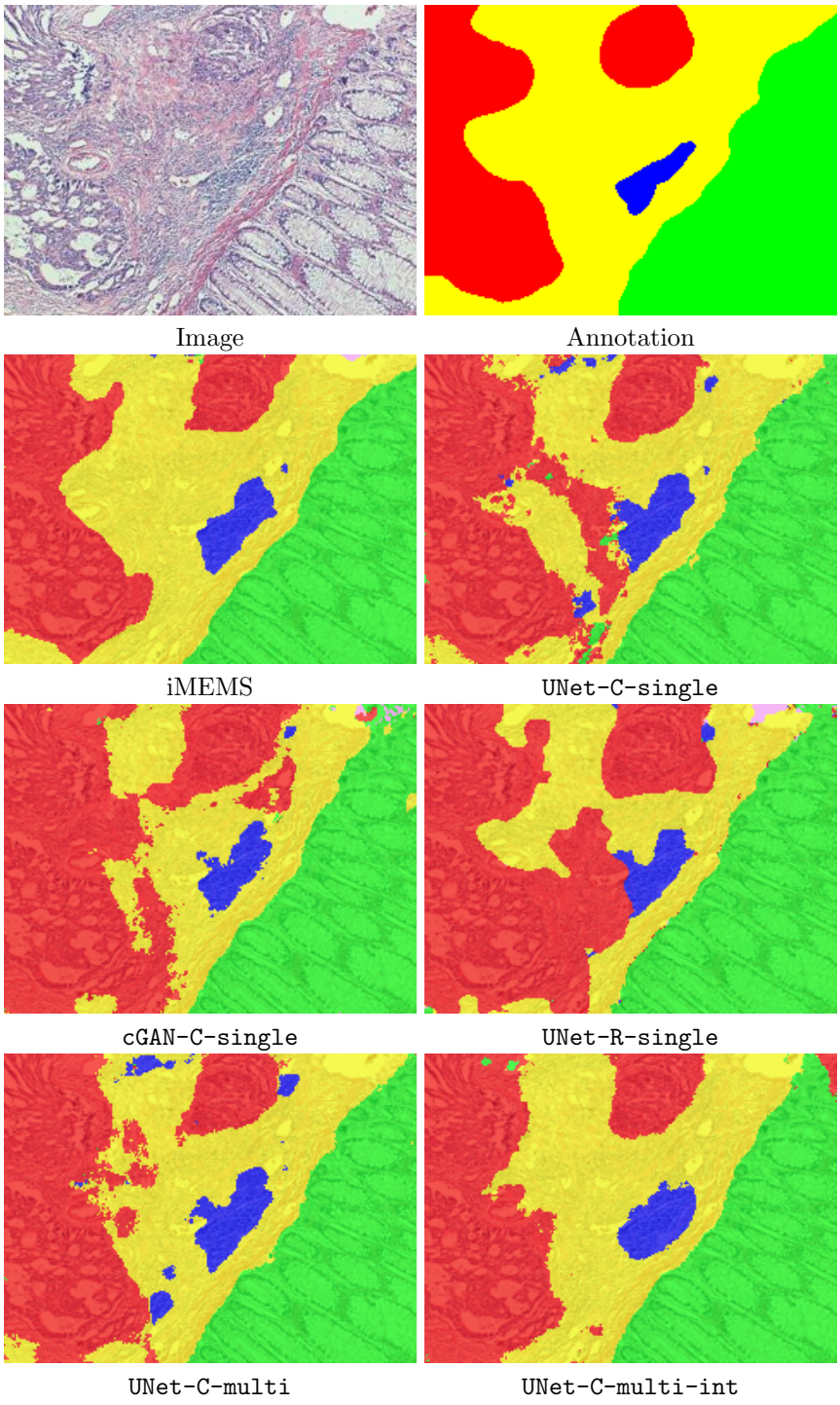


Figure 4.7: For the *in-house colon dataset*, visual results on an example test image. Segmentation labels are shown with green (normal), red (tumorous), yellow (connective tissue), blue (dense lymphoid tissue), and pink (non-tissue). Results are embedded on original images for better visualization.

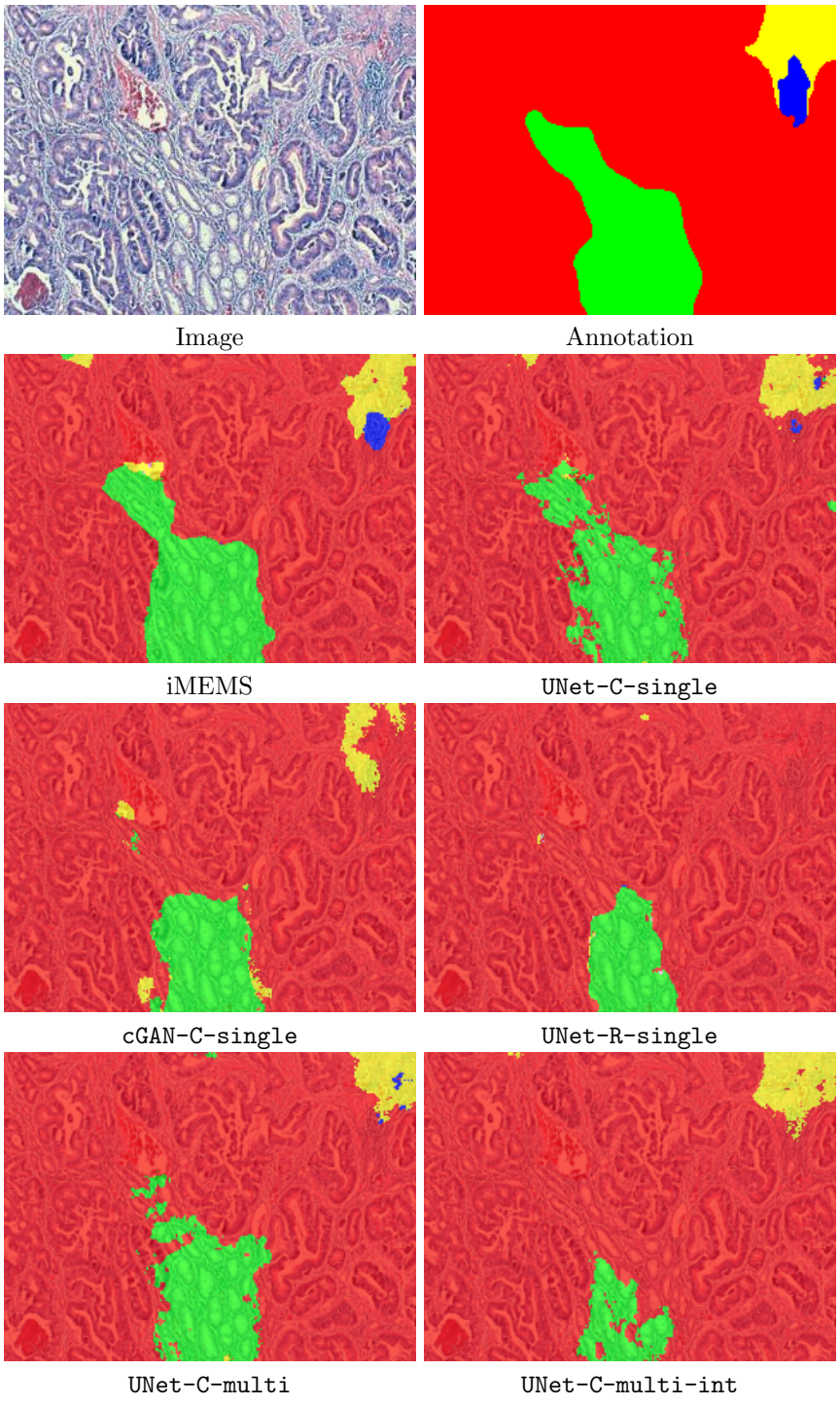


Figure 4.8: For the *in-house colon dataset*, visual results on an example test image. Segmentation labels are shown with green (normal), red (tumorous), yellow (connective tissue), blue (dense lymphoid tissue), and pink (non-tissue). Results are embedded on original images for better visualization.

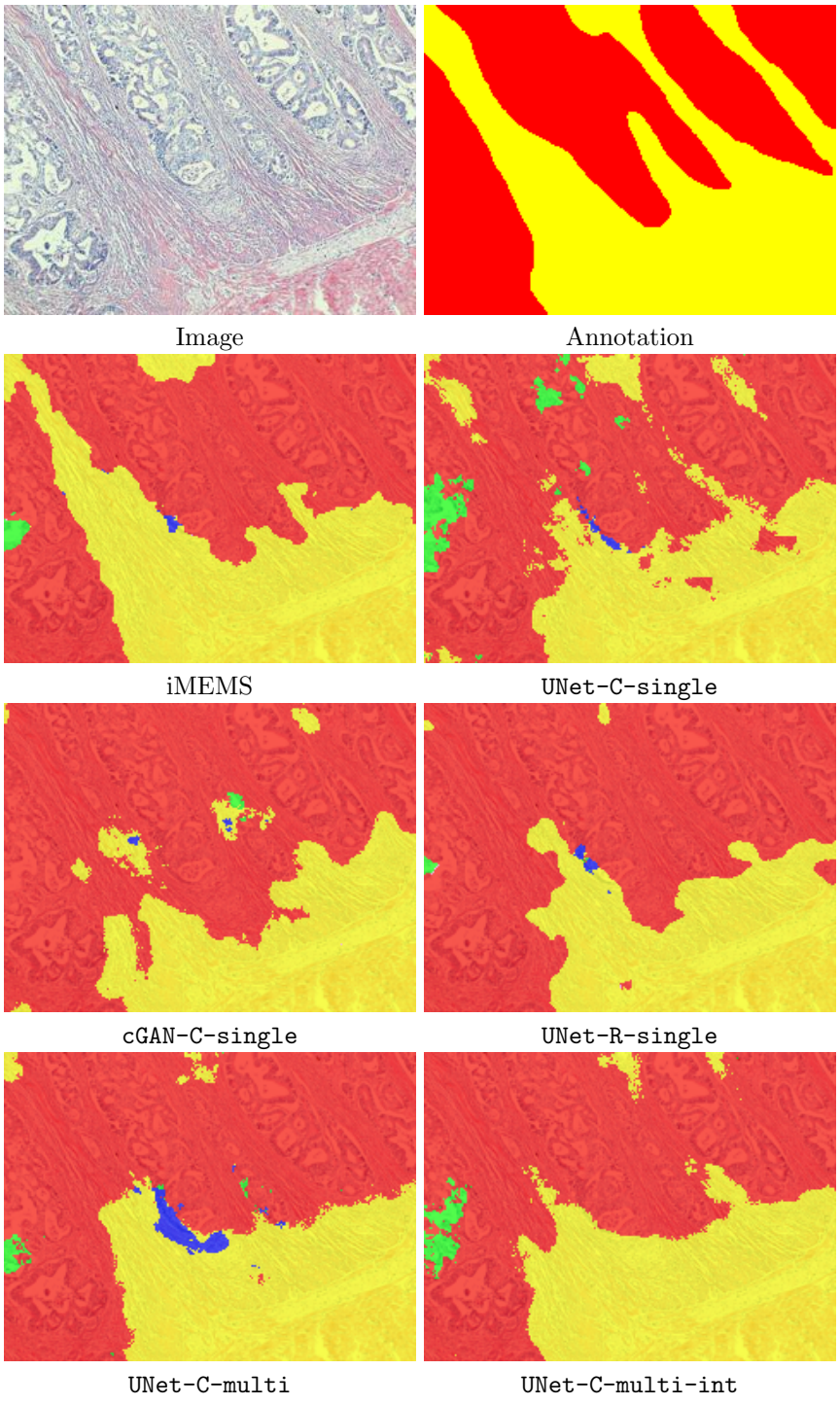


Figure 4.9: For the *in-house colon dataset*, visual results on an example test image. Segmentation labels are shown with green (normal), red (tumorous), yellow (connective tissue), blue (dense lymphoid tissue), and pink (non-tissue). Results are embedded on original images for better visualization.

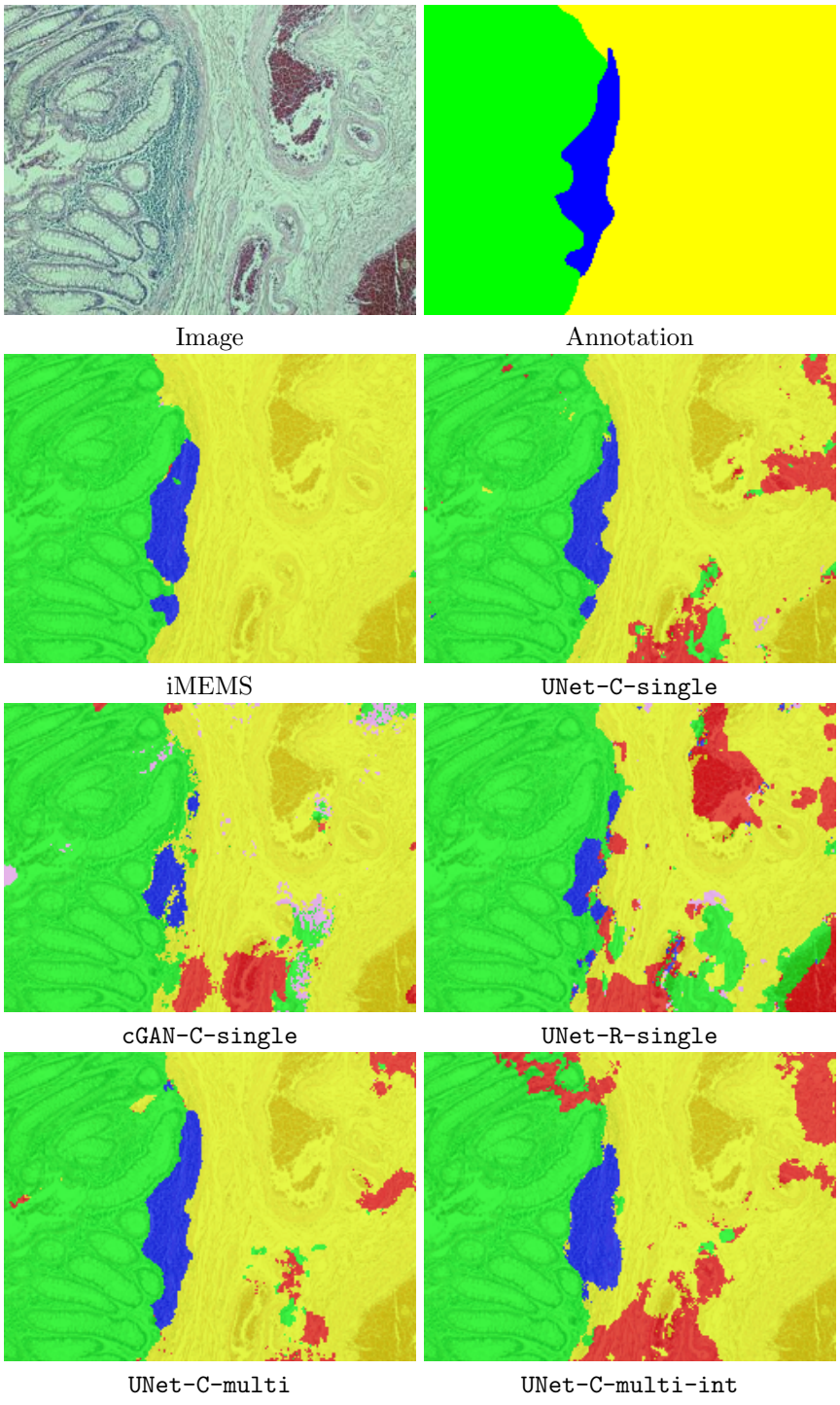


Figure 4.10: For the *in-house colon dataset*, visual results on an example test image. Segmentation labels are shown with green (normal), red (tumorous), yellow (connective tissue), blue (dense lymphoid tissue), and pink (non-tissue). Results are embedded on original images for better visualization.

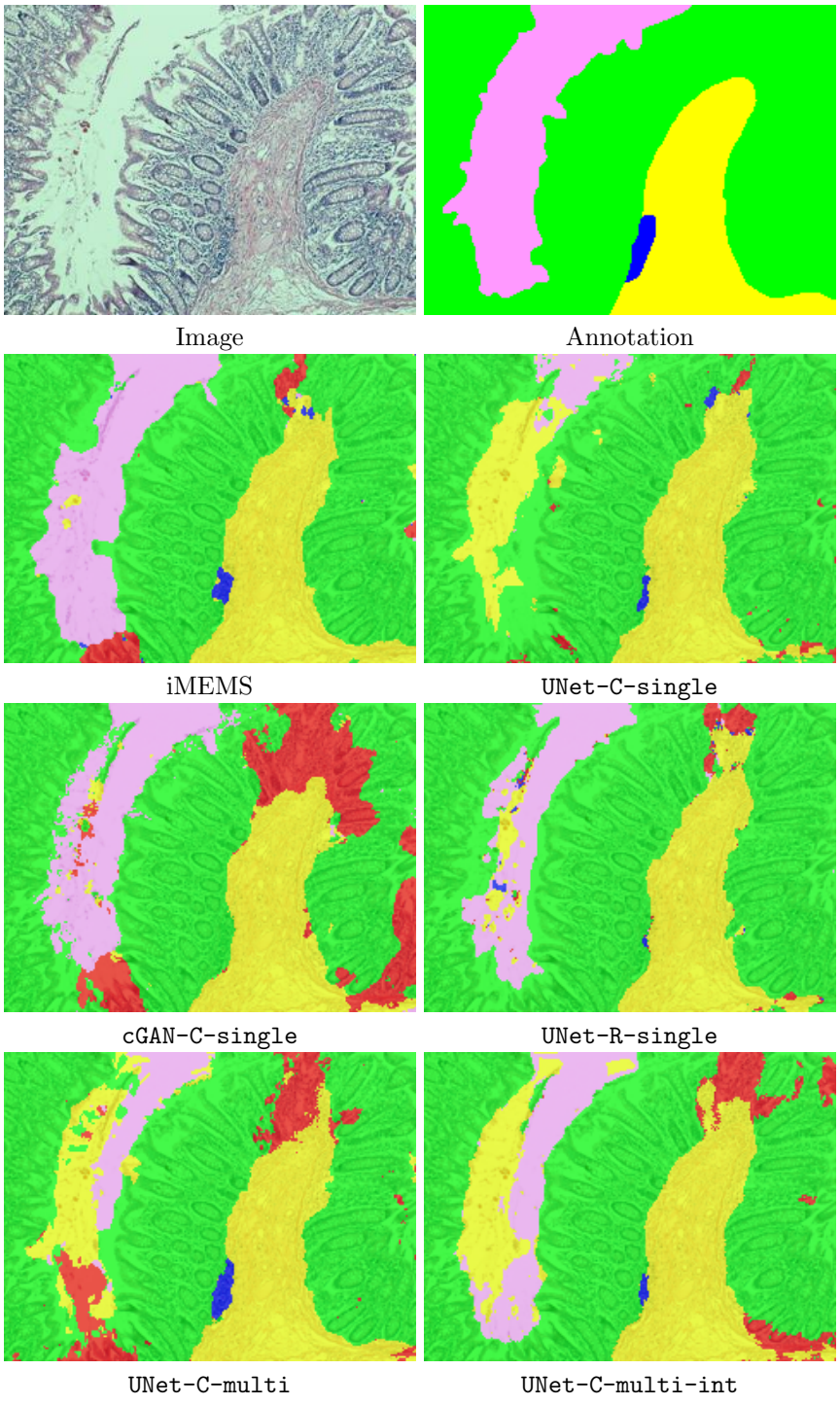


Figure 4.11: For the *in-house colon dataset*, visual results on an example test image. Segmentation labels are shown with green (normal), red (tumorous), yellow (connective tissue), blue (dense lymphoid tissue), and pink (non-tissue). Results are embedded on original images for better visualization.

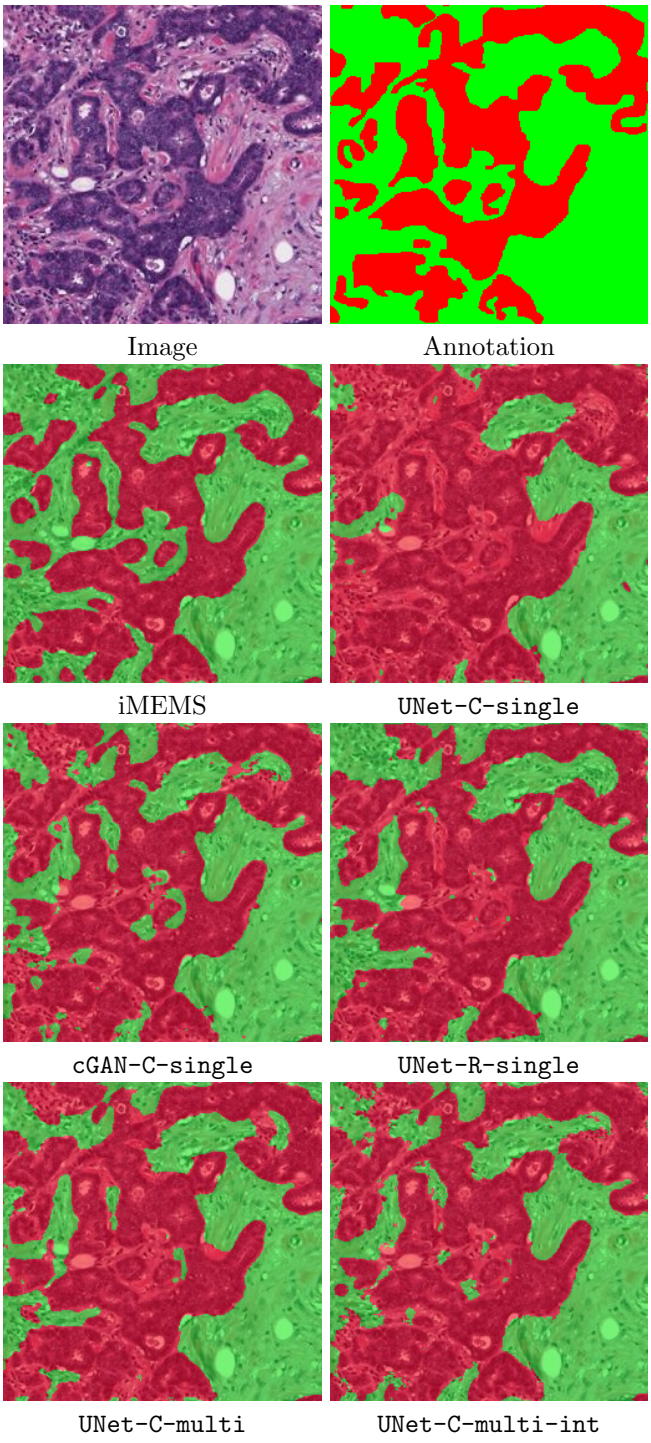


Figure 4.12: For the *epithelium dataset*, visual results on an example test image. Segmentation labels are shown with red (epithelium) and green (background). Results are embedded on original images for better visualization.

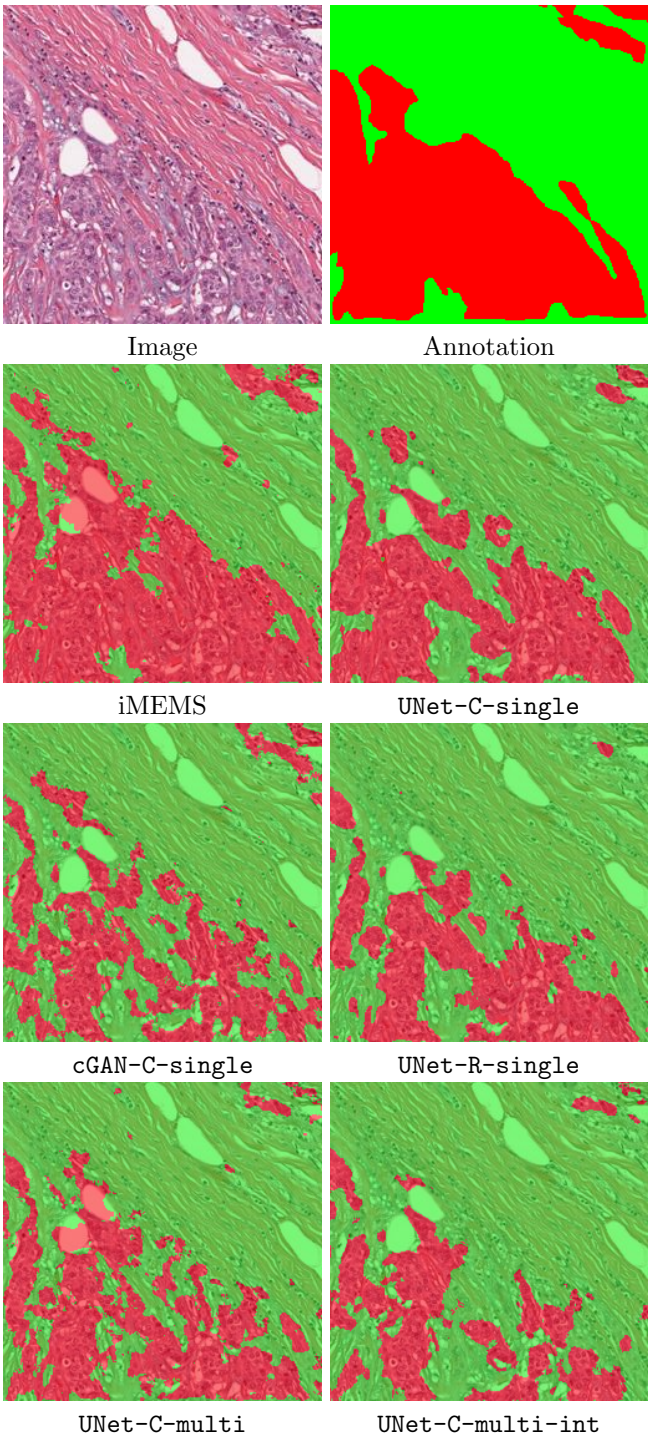


Figure 4.13: For the *epithelium dataset*, visual results on an example test image. Segmentation labels are shown with red (epithelium) and green (background). Results are embedded on original images for better visualization.

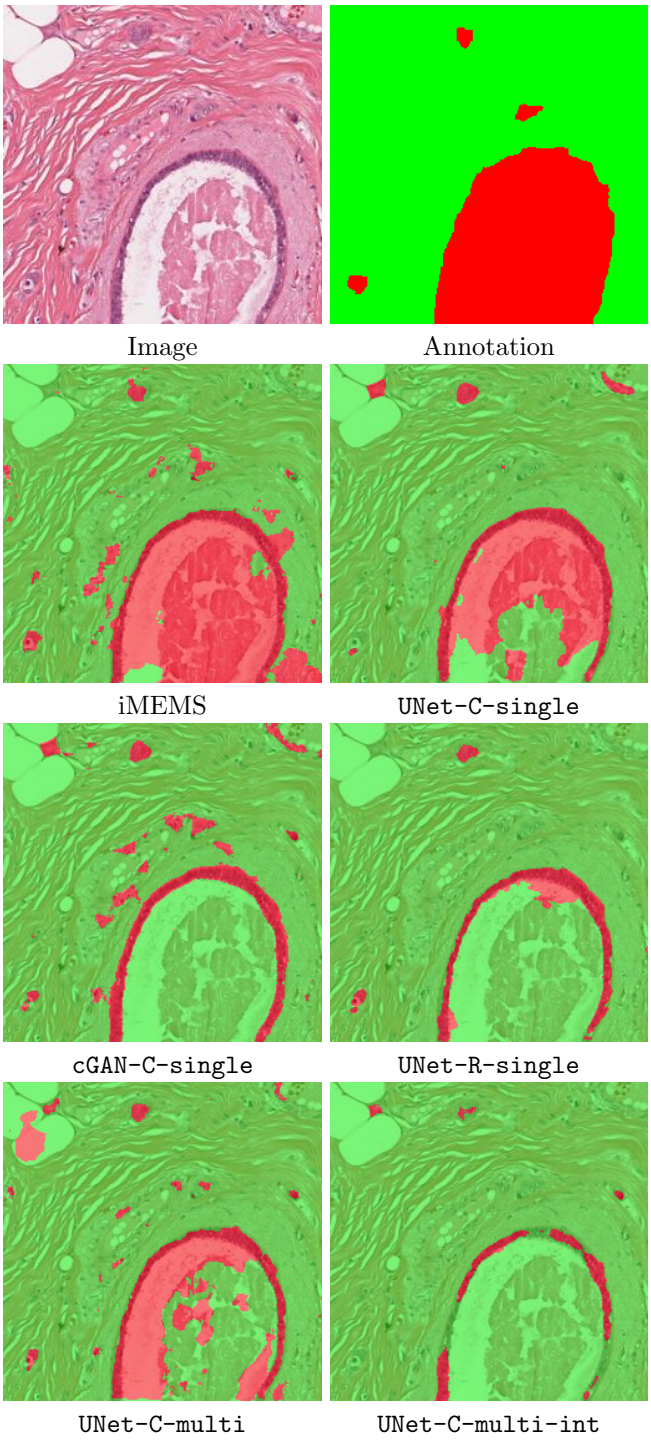


Figure 4.14: For the *epithelium dataset*, visual results on an example test image. Segmentation labels are shown with red (epithelium) and green (background). Results are embedded on original images for better visualization.

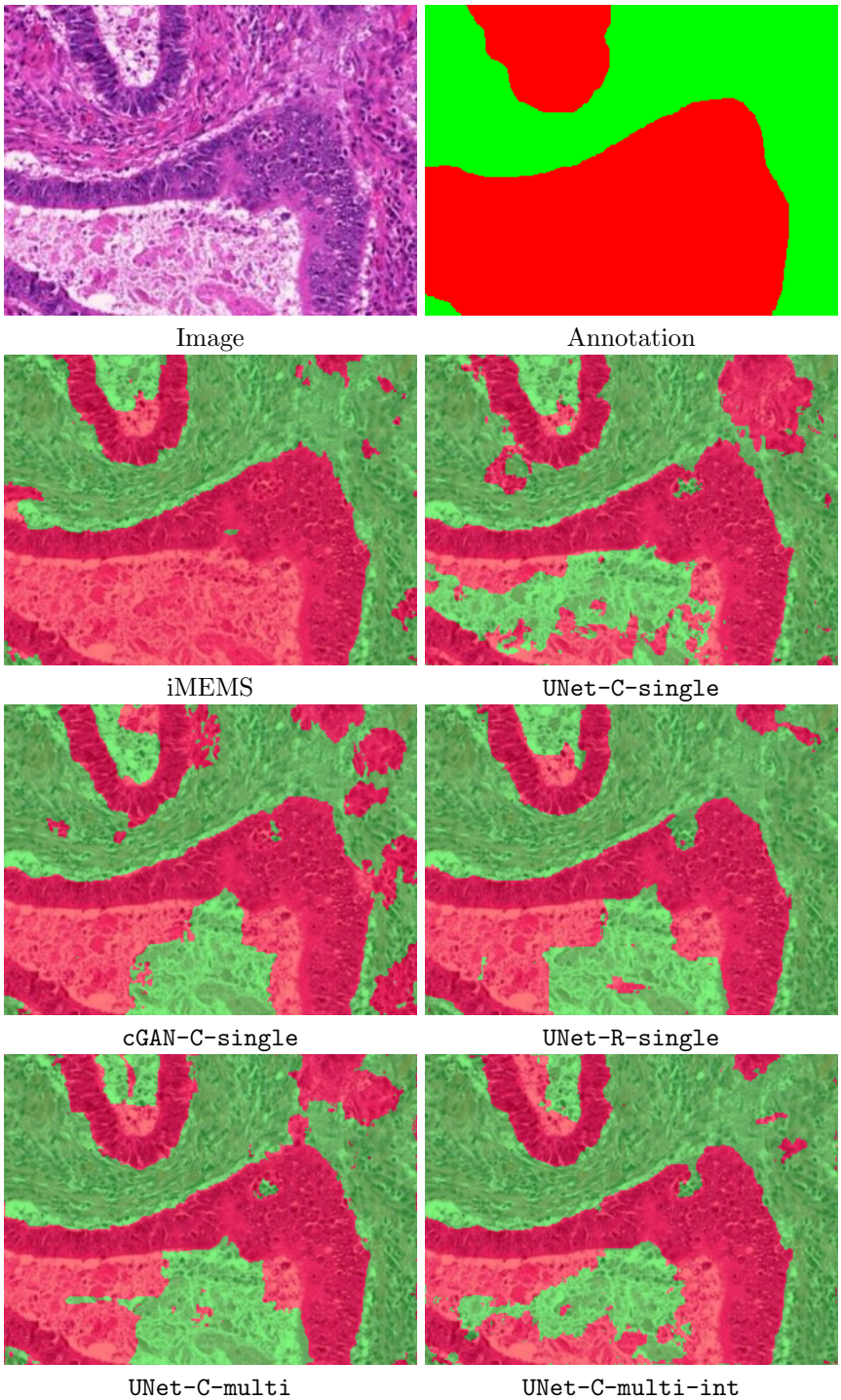


Figure 4.15: For the *tubule dataset*, visual results on an example test image. Segmentation labels are shown with red (tubule) and green (background). Results are embedded on original images for better visualization.

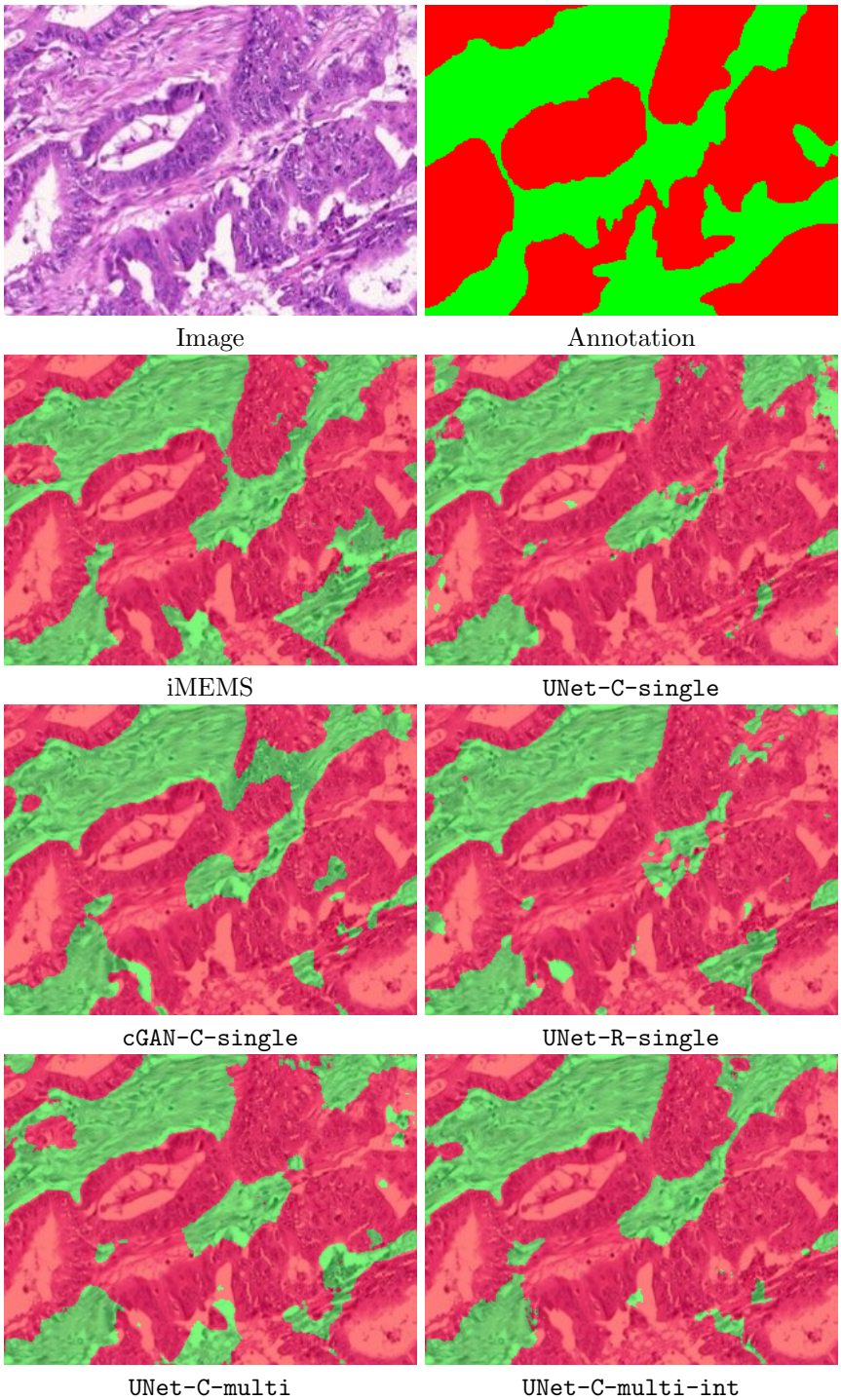


Figure 4.16: For the *tubule dataset*, visual results on an example test image. Segmentation labels are shown with red (tubule) and green (background). Results are embedded on original images for better visualization.

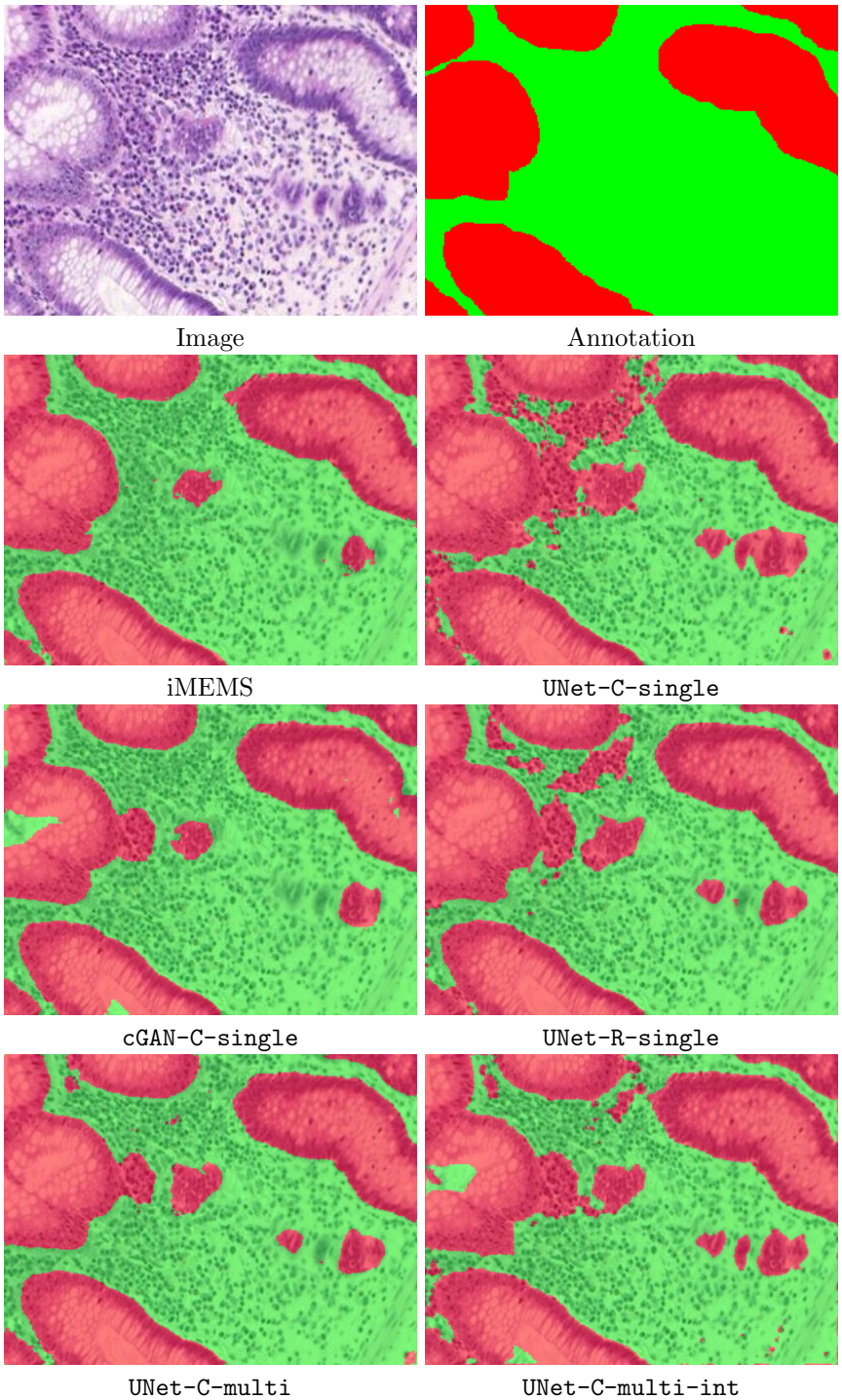


Figure 4.17: For the *tubule dataset*, visual results on an example test image. Segmentation labels are shown with red (tubule) and green (background). Results are embedded on original images for better visualization.

estimates the original segmentation map using a UNet. `cGAN-C-single` estimates the segmentation map but this time with the cGAN also used by iMEMS. `UNet-R-single` also estimates the proposed embedding but not using a cGAN. The results in Table 4.1 show that the contribution of both factors is critical to obtain the best results. Furthermore, they show that the proposed embedding provides effective regularization for network training regardless of the network type. `UNet-R-single` improves the results of `UNet-C-single` and iMEMS improves those of `cGAN-C-single`. Nevertheless, the proposed embedding together with the cGAN yields better improvement.

Next, we compare iMEMS with another regularization technique that simultaneously minimizes supervised and unsupervised losses defined on the segmentation and image reconstruction tasks, respectively. This technique relies on constructing a multi-task network whose weights are learned by minimizing a joint loss function [12, 14]. For the supervised loss, $\mathcal{L}_{seg}$, the average cross-entropy is used. For the unsupervised loss, two definitions are used. First is the reconstruction loss, $\mathcal{L}_{rec}$, defined at the input level; it is the mean square error between the input and reconstructed images. Second is the sum of the reconstruction losses, $\mathcal{L}_{int}$, at the intermediate layers; they are the mean square errors between the maps of the corresponding encoders and decoders. Here two more comparison algorithms are implemented. `UNet-C-multi` linearly combines the supervised loss with the reconstruction loss at the input level without considering those defined at the intermediate layers whereas `UNet-C-multi-int` also considers the latter losses. Here two variants are implemented since it becomes harder to select the right contribution of each loss in the joint loss function as the number of losses increases. These variants are to better understand this phenomenon.

`UNet-C-multi` defines its joint loss function as

$$\mathcal{L}_{model} = \lambda_{seg}\ \mathcal{L}_{seg} + \lambda_{rec}\ \mathcal{L}_{rec} \tag{4.2}$$

where λ_{seg} and λ_{rec} are the coefficients of the supervised and unsupervised losses, respectively. Here to find a good combination of these coefficients, we set $\lambda_{rec} = (1 - \lambda_{seg})$ and perform the grid search on the test set/fold images. In Figures 4.18(a), 4.18(c), and 4.18(e), the metrics are plotted as a function of λ_{seg}

for the in-house colon, epithelium, and tubule datasets, respectively. When λ_{seg} is too small, the performance of the segmentation task decreases dramatically. On the contrary, when it is close to 1, the image reconstruction task cannot help improve the results. This grid search selects $\lambda_{seg} = 0.6$, which gives the best average F-score for the in-house colon dataset. Using the same approach, $\lambda_{seg} = 0.4$ is selected for the other two datasets. Table 4.1 and Figures 4.7-4.17 present the results for these λ_{seg} values. These results show that a multi-task network, which regularizes its training by simultaneously minimizing the supervised and unsupervised losses, improves the results of the single-stage networks. On the other hand, iMEMS leads to better results. The reason might be the following: First, iMEMS unites the supervised and unsupervised tasks into a single one and trains its network by minimizing the loss defined on the united task. This united task provides a very natural way of loss definition, eliminating the necessity of defining a joint loss function with right contributions of the supervised and unsupervised losses. This may provide more effective regularization for employing unsupervised learning in network training. Second, iMEMS learns this united task by benefiting from the well-known synthesizing ability of cGANs. Thanks to using a cGAN, iMEMS produces realistic outputs that better comply with spatial contiguity.

`UNet-C-multi-int` defines a similar loss function, but this time, also considering the sum of the reconstruction losses, $\mathcal{L}_{int}$, at the intermediate layers. It defines the following joint loss function, which is also used in [12, 14] to regularize their network training.

$$\mathcal{L}_{model} = \lambda_{seg}\ \mathcal{L}_{seg} + \lambda_{rec}\ \mathcal{L}_{rec} + \lambda_{int}\ \mathcal{L}_{int} \tag{4.3}$$

As aforementioned, as their number increases, it becomes harder to adjust the coefficients relative to each other. In our experiments, we use the best configuration of $\lambda_{seg} = 0.6$ and $\lambda_{rec} = 0.4$ selected by `UNet-C-multi` for the in-house colon dataset and $\lambda_{seg} = 0.4$ and $\lambda_{rec} = 0.6$ for the other datasets, and determine the coefficient λ_{int} also by the grid search. This grid search gives the best average F-score when λ_{int} is 0.8, 0.7, and 0.3, for the in-house colon, epithelium, and tubule datasets, respectively. For the in-house colon, epithelium, and tubule datasets, the metrics are plotted as a function of λ_{int} in Figures 4.18(b), 4.18(d),

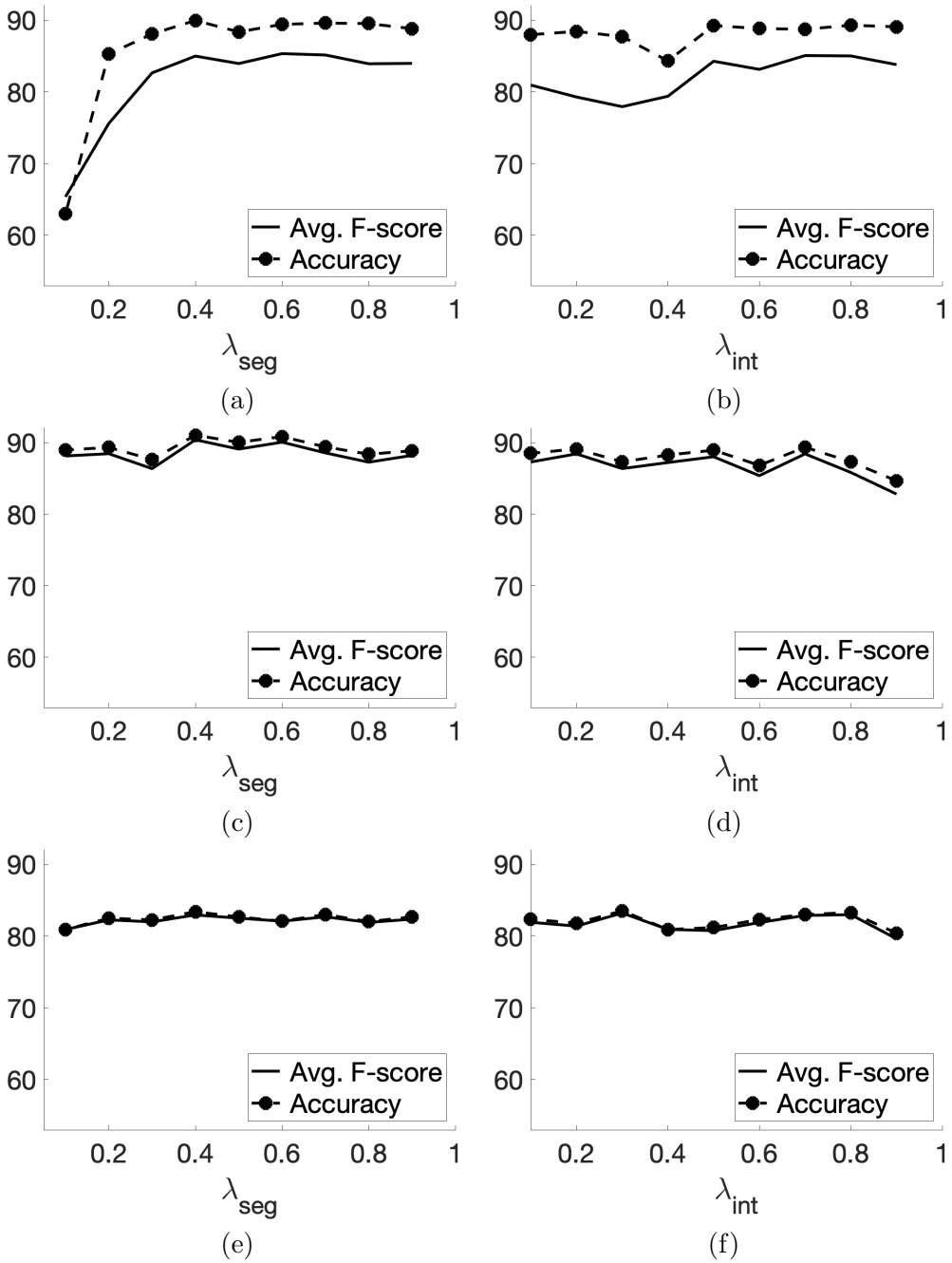


Figure 4.18: Accuracy and average F-scores of `UNet-C-multi` as a function of λ_{seg} (a) for the in-house colon, (c) epithelium, and (e) tubule datasets, respectively. Accuracy and average F-scores of `UNet-C-multi-int` as a function of λ_{int} (b) for the in-house colon, (d) epithelium, and (f) tubule datasets, respectively.

and 4.18(f), respectively. The test set/fold results for these λ_{int} values are provided in Table 4.1. Here it is observed that the inclusion of the intermediate layer losses does not help further improve the results. The reason might be the following: The linear function, which is used by `UNet-C-multi-int` as well as by the previous studies [12, 14], may not be the best way to combine these losses and/or it may require a more thorough coefficient search. On the contrary, the iMEMS method requires neither such an explicit joint loss function definition nor such a coefficient search since its proposed united task intrinsically combines these losses.

The comparison methods presented in quantitative and visual results are designed and implemented in order to evaluate the effectiveness of the proposed contributions. On the other hand, comparing the proposed iMEMS method with recent studies using these publicly available datasets can reveal the holistic contribution of the method to the digital pathology literature. To this end, we compare iMEMS with two deep learning studies using the publicly available epithelium and tubule datasets. The first study [105] employs a fully convolutional residual network (FCRN) followed by a pyramid dilated convolution (PDC) module to obtain multi-level and multi-scale contextual information. The second study [106] proposes to use a recurrent residual convolutional neural network based on UNet architecture (R2U-Net) for semantic segmentation. While the first study reports only the average F-score, the second study reports the pixel-level accuracy in addition to the average F-score.

The FCRN method firstly reports the F-score for their baseline model which does not include the PDC module and the F-score is 0.8831 ± 0.02 for the epithelium dataset and 0.8242 ± 0.03 for the tubule dataset. Our proposed iMEMS method leads to better results for both datasets than the baseline version of the FCRN method. Next, the FCRN method augments the training data in both datasets by applying transformations including rescaling and horizontal and vertical flipping. In this data augmented version, the reported F-score is 0.8981 ± 0.02 for the epithelium dataset and 0.8646 ± 0.03 for the tubule dataset. In this setting, while the iMEMS method obtains better results in the tubule dataset, the FCRN method produces a higher F-score metric. At this point, it should be noted that

this comparison may not be fair, as the iMEMS method has been trained with a more deficient dataset than the comparison method. Lastly, the first study reports the F-scores of the FCRN model including the PDC module trained with the augmented datasets. The reported F-score is 0.9066 ± 0.01 for the epithelium dataset and 0.8950 ± 0.02 for the tubule dataset. For both datasets, the FCRN method including the PDC module gives higher F-scores than the iMEMS method. By comparing the F-scores presented for the first method consecutively, it is revealed that the performance of the FCRN method is increased essentially by augmenting the dataset rather than the additional PDC module. A fairer comparison can be achieved by training the iMEMS method with a similar augmented dataset.

Before applying the R2U-Net to the epithelium and tubule datasets, the authors have cropped non-overlapping patches from original images in order to obtain more samples and employ networks with less number of parameters. The F-score and pixel-level accuracy are 0.9050 and 0.9254 for the epithelium dataset and 0.9013 and 0.9031 for the tubule dataset, respectively. According to these results, the R2U-Net model produces higher F-scores and accuracy values than the proposed iMEMS method for both datasets. It should be noted that the R2U-Net applied semantic segmentation on the cropped patches and the presented metrics were obtained on these cropped patches but not the entire images.

Lastly, in order to evaluate the effectivenes of the iMEMS method for cross-organ segmentation, we segmented images in the epithelium and tubule datasets with the iMEMS models trained on the tubule and epithelium datasets, respectively. To this end, all images in the epithelium (tubule) dataset are first rescaled to be in appropriate dimensions with the networks trained on the tubule (epithelium) dataset. Then, the rescaled images in the epithelium dataset are segmented using each of the five iMEMS models that was trained on each fold in the tubule (epithelium) dataset. The quantitative results in Table 4.3 reveal that the iMEMS method does not produce accurate segmentation maps in this new setting. The reason might be the following: Since the epithelium dataset includes images of breast tissues and the tubule dataset includes images of colon tissues, whose tissue formations are different, the representation learned for one dataset cannot

Table 4.3: F-scores and accuracies of the proposed iMEMS method for cross-organ segmentation. (a) For the epithelium dataset, these are the average metrics obtained on the five test folds trained on the tubule dataset. (c) For the tubule dataset, these are also the average metrics obtained on the five test folds trained on the epithelium dataset.

	F-scores			
	Epithelium	*Backgr.*	*Average*	**Accuracy**
iMEMS (trained on the epithelium dataset)	**85.51**	**92.40**	**88.96**	**90.17**
iMEMS (trained on the tubule dataset)	52.19	79.92	66.05	71.87

(a)

	F-scores			
	Tubule	*Backgr.*	*Average*	**Accuracy**
iMEMS (trained on the tubule dataset)	**87.08**	**87.00**	**87.04**	**87.09**
iMEMS (trained on the epithelium dataset)	60.64	65.39	63.01	63.80

(b)

sufficiently contribute to the segmentation of images in the other dataset.

4.2.3 Discussion

In Figures 4.7-4.17, it is observed that especially the comparison algorithms yield many small segmented regions, which can be easily corrected by post-processing. To understand how this affects the results, two different post-processing approaches are employed for the in-house colon dataset. The analysis is similar for the other two datasets. First, the following simple post-processing algorithm is applied to the segmentation maps obtained from the iMEMS method and comparison algorithms: Starting from the smallest one, each segmented region smaller than an area threshold τ is merged with its smallest adjacent region. This merge continues until there remains no region smaller than τ. The results reported in Table 4.4(a) indicate that this post-processing is effective to increase the performance. However, this increase is similar for all algorithms and does not change the conclusion drawn from the comparative study. Note that this table reports the average F-scores and the accuracies; both metrics show the same trend. The visual results presented in Figures 4.19-4.21 reveal that this simple post-processing is effective in correcting small incorrect regions, but determining the τ threshold inaccurately may cause the small regions that are already correctly segmented to

be assigned to incorrect classes.

Conditional random fields (CRFs) have been frequently employed as a post-processing step to refine the segmentation maps generated by FCNs [27, 79] since they incorporate pair-wise potentials between adjacent image pixels. Thus, a fully connected CRF [107], which considers pair-wise potentials on all pairs of pixels in the image in a tractable algorithm, is used as the second post-processing method for all competing methods. The quantitative results presented in Table 4.4(b) show that using the CRF method as a post-processing step increases average F-scores and accuracies almost evenly for all competing methods. On the other hand, although the use of CRF improves the performance of all competing methods, it has some disadvantages in the context of the histopathological data used. In Figures 4.19(h) and 4.21(h), the CRF method corrects small connective tissue regions that are segmented incorrectly within normal and tumorous regions and annotates them with the correct class label. However, in Figures 4.20(h) and 4.21(h), connective tissue regions between normal and tumorous regions are incorrectly annotated with normal or tumorous classes by the CRF method since connective tissues occupy small regions between the two adjacent normal and tumorous class regions. Therefore, it is possible to say that, for datasets containing small regions between two adjacent large regions annotated with different classes or small regions annotated with a different class than their surrounding regions, CRFs may incorrectly edit the segmentation maps provided by FCN methods.

4.2.4 Refining the iMEMS method with the DeepFeature method

The iMEMS method is proposed to segment homogeneous regions in heterogeneous colon tissue images. Small segmented regions yielded by the iMEMS method can be easily corrected with the simple and CRF post-processing methods presented in the previous section. On the other hand, these post-processing methods may be insufficient for medium-sized and relatively large regions. In order to correct such regions with a more robust method, one may consider to

Table 4.4: (a) For the in-house colon dataset, test set average F-scores and accuracies of the algorithms after a simple post-processing. The results are reported when the area threshold is selected as $\tau = \{5000, 10000, 25000, 50000\}$ pixels and when no post-processing is applied, i.e., when $\tau = 0$. (b) For the in-house colon dataset, test set average F-scores and accuracies of the algorithms after the CRF post-processing. The results are reported when the CRF post-processing is applied and when no post-processing is applied.

(a)

	Average F-scores					**Accuracies**				
	0	5000	10000	25000	50000	0	5000	10000	25000	50000
iMEMS	87.78	88.30	88.48	88.27	87.64	91.76	91.96	92.14	92.20	92.18
`UNet-C-single`	80.70	81.76	82.00	82.19	81.41	89.27	89.99	90.14	90.50	90.44
`cGAN-C-single`	83.81	84.90	85.20	85.07	84.83	88.49	89.08	89.33	89.55	89.71
`UNet-R-single`	82.22	83.82	84.26	83.96	82.96	88.80	89.31	89.50	89.64	89.51
`UNet-C-multi`	84.70	85.88	86.48	86.30	85.20	89.91	90.56	90.87	91.07	90.95
`UNet-C-multi-int`	84.59	84.91	85.16	85.21	84.59	88.03	88.46	88.73	89.11	89.51

(b)

	Average F-scores		**Accuracies**	
	w/o CRF	w/ CRF	w/o CRF	w/ CRF
iMEMS	87.78	88.73	91.76	92.28
`UNet-C-single`	80.70	81.51	89.27	90.29
`cGAN-C-single`	83.81	85.26	88.49	89.86
`UNet-R-single`	82.22	83.63	88.80	89.49
`UNet-C-multi`	84.70	86.81	89.91	91.28
`UNet-C-multi-int`	84.59	85.69	88.03	89.48

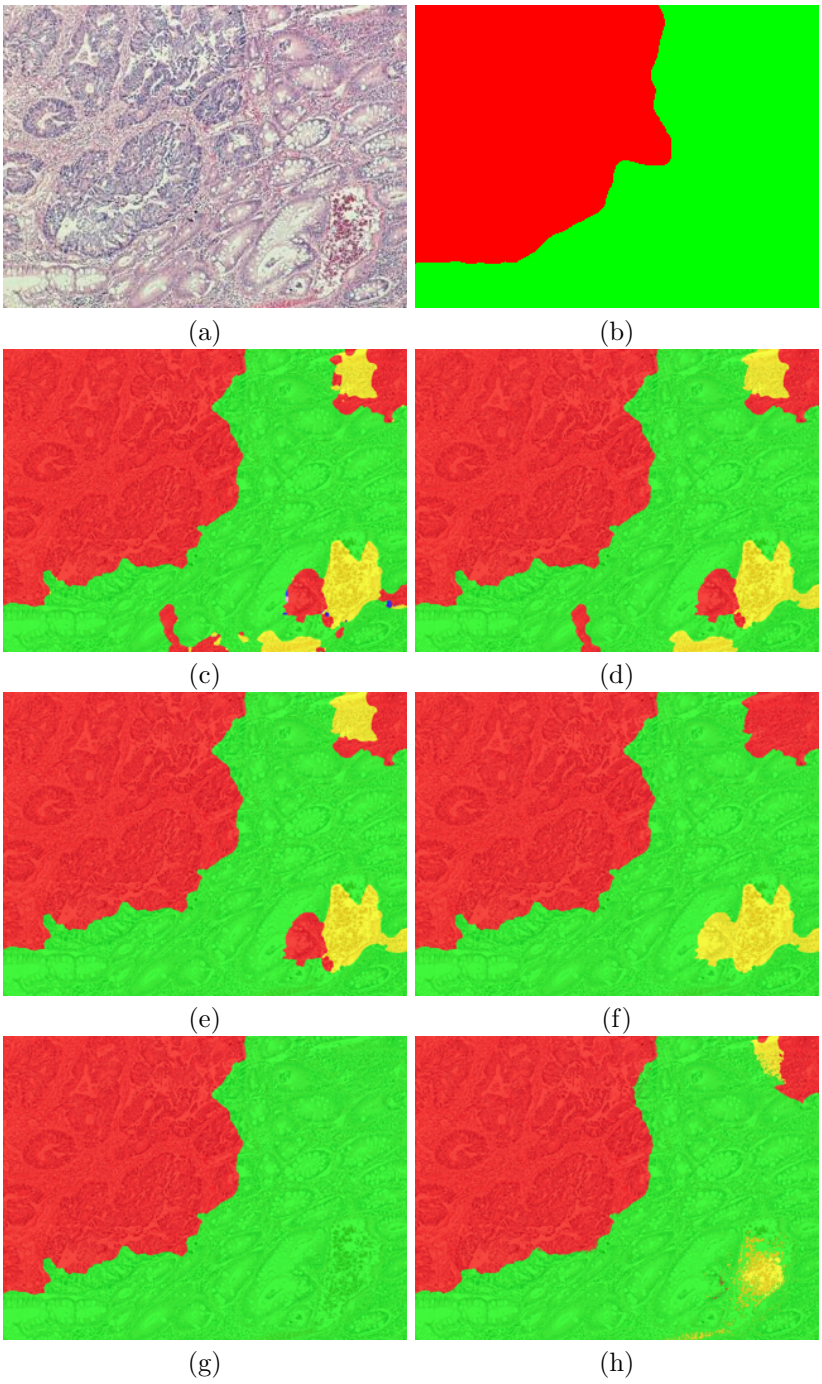


Figure 4.19: For the *in-house colon dataset*, visual results on an example test image after applying post-processing. (a) An original input image. (b) Its ground truth segmentation map. (c) Segmentation map generated by iMEMS method. Segmentation maps after applying the simple post-processing with different area thresholds τ; (d) $\tau = 5000$, (e) $\tau = 10000$, (f) $\tau = 25000$, and (g) $\tau = 50000$ pixels. (h) Segmentation map after applying the CRF post-processing.

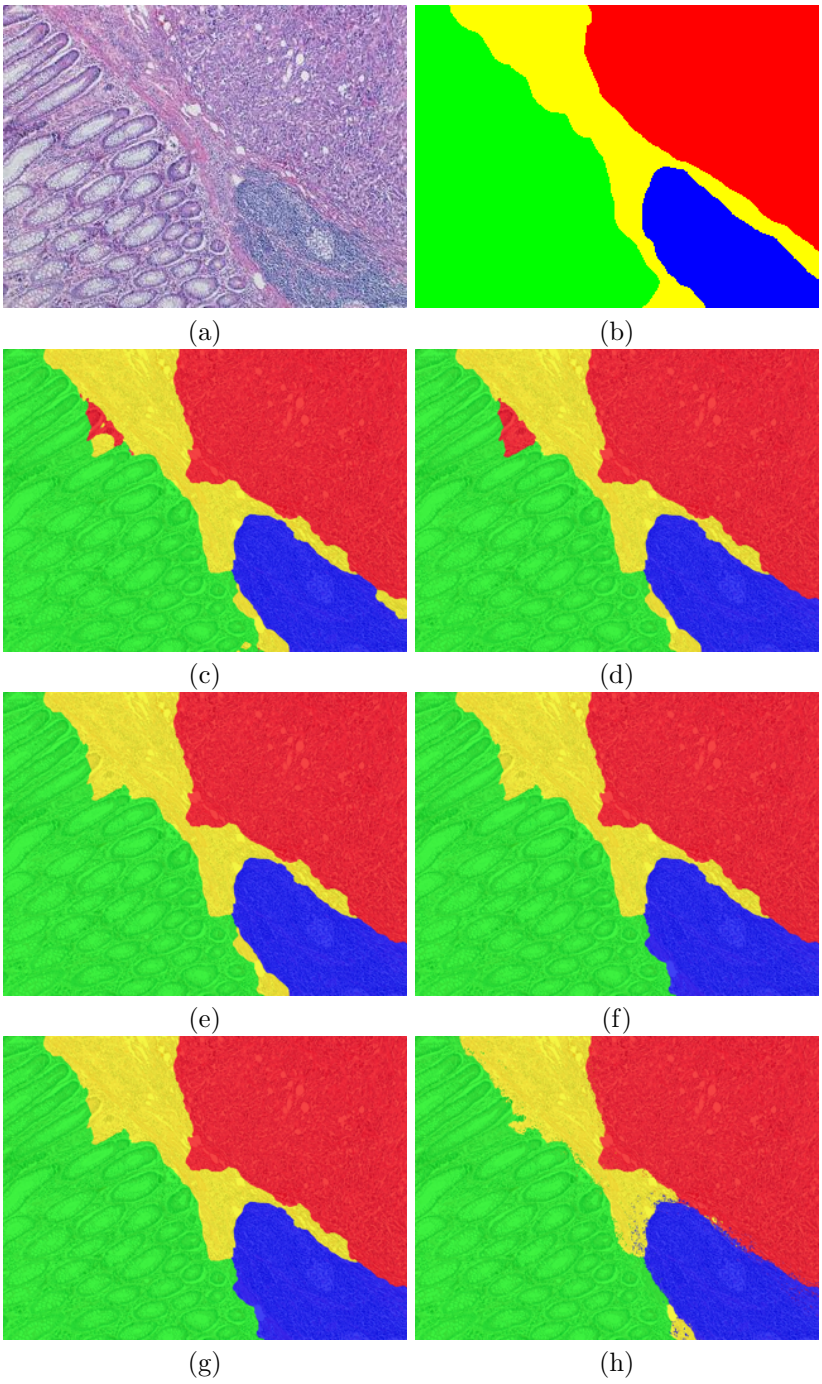


Figure 4.20: For the *in-house colon dataset*, visual results on an example test image after applying post-processing. (a) An original input image. (b) Its ground truth segmentation map. (c) Segmentation map generated by iMEMS method. Segmentation maps after applying the simple post-processing with different area thresholds τ; (d) $\tau = 5000$, (e) $\tau = 10000$, (f) $\tau = 25000$, and (g) $\tau = 50000$ pixels. (h) Segmentation map after applying the CRF post-processing.

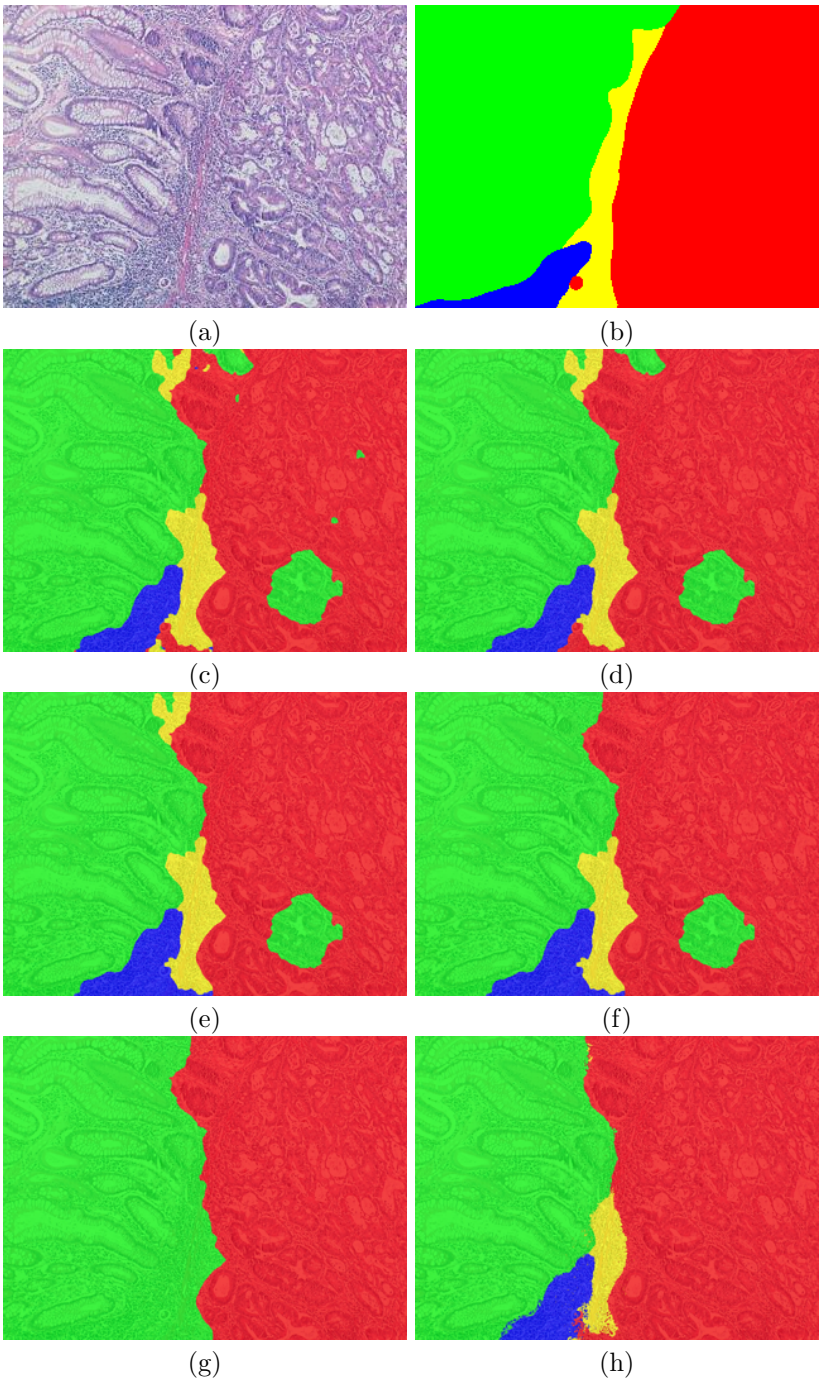


Figure 4.21: For the *in-house colon dataset*, visual results on an example test image after applying post-processing. (a) An original input image. (b) Its ground truth segmentation map. (c) Segmentation map generated by iMEMS method. Segmentation maps after applying the simple post-processing with different area thresholds τ; (d) $\tau = 5000$, (e) $\tau = 10000$, (f) $\tau = 25000$, and (g) $\tau = 50000$ pixels. (h) Segmentation map after applying the CRF post-processing.

incorporate the *DeepFeature* method into the iMEMS method. To explore this possibility, we have designed the following postprocessing method and applied it to the results obtained on the in-house colon dataset. For that, first, salient subregions are defined around cytological tissue components within the training and test images of this dataset. Then, these subregions are characterized by learning their local features using a DBN in an unsupervised way. These local features are clustered by the k-means algorithm and each subregion is represented with a cluster label with respect to its local feature vector. Unlike the homogeneous dataset used in the *DeepFeature* study, the training and test images in the in-house colon dataset contain regions belonging to different classes. Therefore, to train the SVM classifier, a histogram of the subregion labels is calculated for each region in the ground truth maps of the training images.

With the aforementioned approach, 266 regions of five classes are obtained from 80 tissue images in the training set of the in-house colon dataset. Then, for each test image, the segmentation map is estimated by the iMEMS method and each segmented region R in this map is refined by this trained SVM classifier. To this end, the salient subregions in each segmented region R are located, their deep local features are calculated by the DBN and their cluster labels are found based on these local features. Then, the histogram for region R is calculated on these cluster labels and inputted to the SVM classifier. The output of the SVM classifier is used as the new semantic label of this region R.

The visual results obtained by the proposed approach are presented in Figure 4.22. First of all, it should be noted that not only medium-sized regions but also large regions within the test images are reclassified by the *DeepFeature* method. According to the visual results, the *DeepFeature* method has also reclassified large regions that were correctly classified by the iMEMS method in the same way as successful. This reveals that the *DeepFeature* method gives accurate results in relatively large regions containing sufficiently large number of salient subregions, and hence, cluster labels. In the meantime, it is observed that incorrectly segmented medium-sized regions, which represent the main purpose of the proposed approach, are refined with the proposed approach. Medium-sized regions that were incorrectly segmented by the iMEMS method within the large

regions in Figures 4.22(a-d) are refined with the *DeepFeature* method and classified with the correct class label. In Figure 4.22(e), the region annotated as empty by the iMEMS method is incorrectly classified as connective tissue by the *DeepFeature* method. The reason might be the following: Regions identified as empty are frequently occurred within regions belonging to different classes, and in these cases, they are annotated with the class label surrounding them. Since the *DeepFeature* method is trained on homogeneous regions independent of each other, it does not take into account the context and neighborhood information of the image and can make such erroneous classifications. Finally, it should be noted that the small-sized regions do not contain sufficiently large number of salient subregions, and hence, cluster labels, so the classification of these regions by the *DeepFeature* method cannot produce reliable results. In these cases, it would be more appropriate to use the annotations estimated by the iMEMS method.

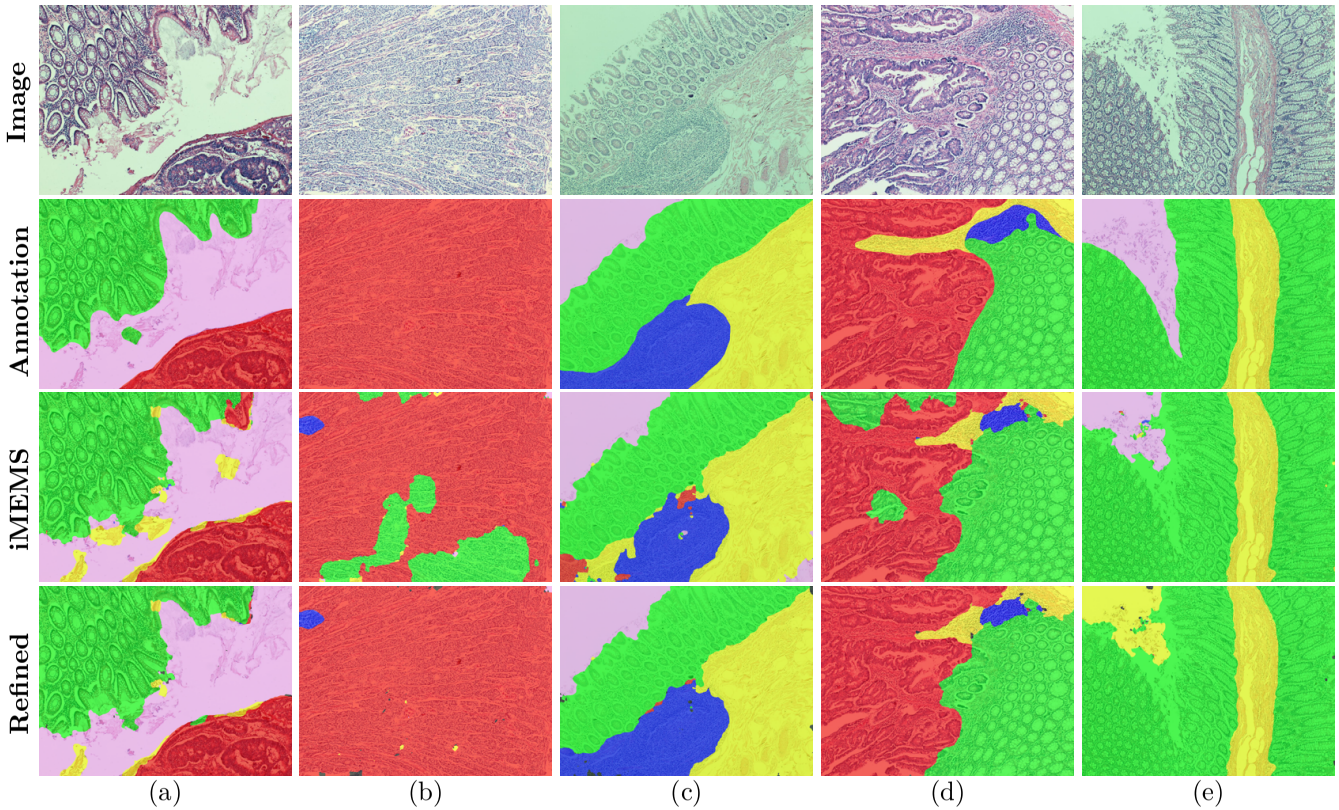


Figure 4.22: For the *in-house colon dataset*, visual results refined by *DeepFeature* method on example test images. Segmentation labels are shown with green (normal), red (tumorous), yellow (connective tissue), blue (dense lymphoid tissue), and pink (non-tissue). Results are embedded on original images for better visualization.

Chapter 5

Conclusion

Digital pathology aims to provide auxiliary tools for pathology in addition to manual examination of histopathological images by expert pathologists in order to prevent the error-prone human factor and overlong examination periods. Advances in artificial intelligence and machine learning lead fast and accurate methods to be used in digital pathology systems. Traditional machine learning methods aim to perform histopathological image analysis with the handcrafted features they define, but the performance of these methods is also directly dependent on the quality of these handcrafted features, and hence, how these features are defined. Deep learning methods, which are frequently employed in many fields recently, extract these features from data directly. However, many deep learning models proposed in the field of histopathological image analysis require annotated data that are limited and difficult to obtain. To address these shortcomings, this thesis introduces deep learning approaches for histopathological image analysis to learn features directly from data instead of using handcrafted features and incorporates unsupervised learning into the supervised objectives to avoid the inadequacy of annotated data. In this regard, it introduces two deep learning methods for the classification and segmentation of histopathological images by exploiting unsupervised learning for feature extraction and training regularization purposes.

The first study presents a semi-supervised classification method for histopathological tissue images. As its first contribution, this method proposes to determine salient subregions in an image and to use only the quantizations (characterizations) of these salient subregions for image representation and classification. As the second contribution, it introduces a new unsupervised technique to learn the subregion quantizations. For that, it proposes to construct a deep belief network of consecutive RBMs whose first layer takes the pixels of a salient subregion and to define the activation values of the hidden unit nodes in the final RBM as its deep features. It then feeds these deep features to a clustering algorithm for learning the quantizations of the salient subregions in an unsupervised way. As its last contribution, this study is a successful demonstration of using restricted Boltzmann machines in the domain of histopathological image analysis. We tested our method on two datasets of microscopic histopathological images of colon tissues. Our experiments revealed that characterizing the salient subregions by the proposed local deep features and using the distribution of these characterized subregions for tissue image representation lead to more accurate classification results compared to the existing algorithms.

The second study proposed the iMEMS method that employs unsupervised learning to regularize the training of a fully convolutional network for a supervised task. This method proposes to define a new embedding to unite the main supervised task of semantic segmentation and an auxiliary unsupervised task of image reconstruction into a single task and to learn this united task by a conditional generative adversarial network. Since the proposed embedding corresponds to a segmentation map that preserves a reconstructive ability, the united task of its learning enforces the network to jointly learn image features and context features. This joint learning lends itself to more effective regularization, leading to better segmentation results. Additionally, this united task provides an intrinsic way of combining the segmentation and image reconstruction losses. Thus, it attends to the difficulty of defining an effective joint loss function to combine the separately defined segmentation and image reconstruction losses in a balanced way. We tested this method for semantic tissue segmentation on three datasets

of histopathological images. Our experiments revealed that it leads to more accurate results compared to its counterparts.

5.1 Future Work

The *DeepFeature* method uses the histogram of quantized salient subregions for defining a global feature set for the entire image. One future research direction is to investigate the other ways of defining this global feature set, such as defining texture measures on the quantized subregions. Another research direction is to explore the use of different network architectures. For example, one may consider combining the activation values in different hidden layers to define a new set of deep features. On an example application, we have discussed how the proposed system can be used in a digital pathology setup. The design of sophisticated algorithms for this purpose is another future research direction of this study.

The iMEMS method is proposed to segment a heterogeneous tissue image into its homogeneous regions. Thus, it can be easily applied to segmenting tissue compartments in whole slide images (WSIs), as in the case of many previous studies. To do so, a WSI can be divided into image tiles, on which the method predicts the output. Alternatively, an image window can be slid on the WSI and the estimated outputs can be averaged to obtain the final segmentation. This application can be considered as one future research direction. The focus of this study is to segment a histopathological image into its tissue compartments. It is possible to extend this idea for the instance segmentation problem in histopathological images. This extension may require modifying the embedding such that it also covers additional supervised tasks (such as the task of predicting instance boundaries) that might be important for instance segmentation. The investigation of this possibility is considered as another future research direction.

We also discuss how to incorporate both methods into a single one. For this purpose, the *DeepFeature* method is used to refine the regions in the segmentation maps obtained by the iMEMS method since *DeepFeature* method is quite

accurate in classifying homogeneous regions. On the other hand, as mentioned in the previous subsection, the success of the *DeepFeature* method is directly proportional to the size of the regions segmented by the iMEMS method and it may not produce accurate classifications for small-sized regions. Therefore, in a future work, the class label of one of the two methods can be chosen according to the size of the segmented regions and the amount of salient subregions it contains. In addition to this, in order to obtain a sufficiently large number of cluster labels, the *DeepFeature* method can classify fixed-size windows in a sliding window scheme instead of regions segmented by the iMEMS method. Then, final class labels of pixels can be obtained by combining these classifications with a method similar to majority voting.

Bibliography

[1] L. Pantanowitz, "Digital images and the future of digital pathology," *Journal of Pathology Informatics*, vol. 1, 2010.

[2] S. Al-Janabi, A. Huisman, and P. J. Van Diest, "Digital pathology: current status and future perspectives," *Histopathology*, vol. 61, no. 1, pp. 1–9, 2012.

[3] F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," *Annual Review of Pathology: Mechanisms of Disease*, vol. 8, pp. 331–359, 2013.

[4] A. Madabhushi, "Digital pathology image analysis: opportunities and challenges," *Imaging in Medicine*, vol. 1, no. 1, p. 7, 2009.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] J. Xu, L. Xiang, R. Hang, and J. Wu, "Stacked sparse autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology," in *11th International Symposium on Biomedical Imaging*, pp. 999–1002, IEEE, 2014.

[7] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119–130, 2015.

[8] A. A. Cruz-Roa, J. E. A. Ovalle, A. Madabhushi, and F. A. G. Osorio, "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 403–410, Springer, 2013.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[10] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.

[11] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?," *Journal of Machine Learning Research*, vol. 11, no. Feb, pp. 625–660, 2010.

[12] J. Zhao, M. Mathieu, R. Goroshin, and Y. Lecun, "Stacked what-where auto-encoders," *arXiv preprint arXiv:1506.02351*, 2015.

[13] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Advances in Neural Information Processing Systems*, pp. 3546–3554, 2015.

[14] Y. Zhang, K. Lee, and H. Lee, "Augmenting supervised neural networks with unsupervised objectives for large-scale image classification," in *International Conference on Machine Learning*, pp. 612–621, 2016.

[15] T. Robert, N. Thome, and M. Cord, "Hybridnet: Classification and reconstruction cooperation for semi-supervised learning," in *Proceedings of the European Conference on Computer Vision*, pp. 153–169, 2018.

[16] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.

[17] O. Jimenez-del Toro, S. Otálora, M. Andersson, K. Eurén, M. Hedlund, M. Rousson, H. Müller, and M. Atzori, "Analysis of histopathology images: From traditional machine learning to deep learning," in *Biomedical Texture Analysis*, pp. 281–314, Elsevier, 2017.

[18] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[19] Y. Bengio, Y. LeCun, *et al.*, "Scaling learning algorithms towards AI," *Large-Scale Kernel Machines*, vol. 34, no. 5, pp. 1–41, 2007.

[20] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[21] C. T. Sari and C. Gunduz-Demir, "Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images," *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1139–1149, 2019.

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

[23] J. Arevalo, A. Cruz-Roa, V. Arias, E. Romero, and F. A. González, "An unsupervised feature learning framework for basal cell carcinoma image analysis," *Artificial Intelligence in Medicine*, vol. 64, no. 2, pp. 131–145, 2015.

[24] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 411–418, Springer, 2013.

[25] T. Chen and C. Chefd'Hotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in *International Workshop on Machine Learning in Medical Imaging*, pp. 17–24, Springer, 2014.

[26] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal

regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.

[27] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *arXiv preprint arXiv:1412.7062*, 2014.

[28] M. Naghavi, A. A. Abajobir, C. Abbafati, K. M. Abbas, F. Abd-Allah, S. F. Abera, V. Aboyans, O. Adetokunboh, A. Afshin, A. Agrawal, *et al.*, "Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet*, vol. 390, no. 10100, pp. 1151–1210, 2017.

[29] P. Favoriti, G. Carbone, M. Greco, F. Pirozzi, R. E. M. Pirozzi, and F. Corcione, "Worldwide burden of colorectal cancer: a review," *Updates in Surgery*, vol. 68, no. 1, pp. 7–11, 2016.

[30] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1366–1378, 2007.

[31] R. Rahmadwati, G. Naghdy, M. Ros, and C. Todd, "Computer aided decision support system for cervical cancer classification," in *Applications of Digital Image Processing XXXV*, vol. 8499, p. 849919, International Society for Optics and Photonics, 2012.

[32] F. Bunyak, A. Hafiane, and K. Palaniappan, "Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets," in *Software Tools and Algorithms for Biological Systems*, pp. 413–424, Springer, 2011.

[33] A. N. Esgiar, R. N. Naguib, B. S. Sharif, M. K. Bennett, and A. Murray, "Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa," *IEEE Transactions on Information Technology in Biomedicine*, vol. 2, no. 3, pp. 197–203, 1998.

[34] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi, "A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1205–1218, 2010.

[35] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 697–704, 2003.

[36] A. N. Esgiar, R. N. Naguib, B. S. Sharif, M. K. Bennett, and A. Murray, "Fractal analysis in the detection of colonic cancer images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 1, pp. 54–58, 2002.

[37] P.-W. Huang and C.-H. Lee, "Automatic classification for pathological prostate images based on fractal analysis," *IEEE Transactions on Medical Imaging*, vol. 28, no. 7, pp. 1037–1050, 2009.

[38] H. Qureshi, O. Sertel, N. Rajpoot, R. Wilson, and M. Gurcan, "Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 196–204, Springer, 2008.

[39] O. Sertel, J. Kong, H. Shimada, U. V. Catalyurek, J. H. Saltz, and M. N. Gurcan, "Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development," *Pattern Recognition*, vol. 42, no. 6, pp. 1093–1103, 2009.

[40] Y. Zhang, B. Zhang, F. Coenen, and W. Lu, "Breast cancer diagnosis from biopsy images with highly reliable random subspace classifier ensembles," *Machine Vision and Applications*, vol. 24, no. 7, pp. 1405–1420, 2013.

[41] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in HER2+

breast cancer histopathology," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 642–653, 2009.

[42] C. Demir, S. H. Gultekin, and B. Yener, "Learning the topological properties of brain tumors," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 3, pp. 262–270, 2005.

[43] B. Weyn, G. van de Wouwer, S. Kumar-Singh, A. van Daele, P. Scheunders, E. Van Marck, and W. Jacob, "Computer-assisted differential diagnosis of malignant mesothelioma based on syntactic structure analysis," *Cytometry: The Journal of the International Society for Analytical Cytology*, vol. 35, no. 1, pp. 23–29, 1999.

[44] G. Olgun, C. Sokmensuer, and C. Gunduz-Demir, "Local object patterns for the representation and classification of colon tissue images," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1390–1396, 2013.

[45] D. Altunbay, C. Cigir, C. Sokmensuer, and C. Gunduz-Demir, "Color graphs for automated cancer diagnosis and grading," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 665–674, 2009.

[46] E. Ozdemir and C. Gunduz-Demir, "A hybrid classification model for digital pathology using structural and statistical pattern recognition," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 474–483, 2012.

[47] H. Rezaeilouyeh, A. Mollahosseini, and M. H. Mahoor, "Microscopic medical image classification framework via deep learning and shearlet transform," *Journal of Medical Imaging*, vol. 3, no. 4, p. 044501, 2016.

[48] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," in *23rd International Conference on Pattern Recognition*, pp. 2440–2445, IEEE, 2016.

[49] J. Xie, R. Liu, J. Luttrell IV, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Frontiers in Genetics*, vol. 10, p. 80, 2019.

[50] B. E. Bejnordi, G. Zuidhof, M. Balkenhol, M. Hermsen, P. Bult, B. van Ginneken, N. Karssemeijer, G. Litjens, and J. van der Laak, "Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images," *Journal of Medical Imaging*, vol. 4, no. 4, p. 044504, 2017.

[51] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours," *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[52] S. Manivannan, W. Li, J. Zhang, E. Trucco, and S. J. McKenna, "Structure prediction for gland segmentation with hand-crafted and deep convolutional features," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 210–221, 2017.

[53] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, I. Eric, and C. Chang, "Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features," *BMC Bioinformatics*, vol. 18, no. 1, p. 281, 2017.

[54] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PloS One*, vol. 14, no. 3, p. e0214587, 2019.

[55] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.

[56] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 2016.

[57] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. 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.

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

[59] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *14th International Symposium on Biomedical Imaging*, pp. 933–936, IEEE, 2017.

[60] H. Wang, A. C. Roa, A. N. Basavanhally, H. L. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features," *Journal of Medical Imaging*, vol. 1, no. 3, p. 034003, 2014.

[61] D. A. Van Valen, T. Kudo, K. M. Lane, D. N. Macklin, N. T. Quach, M. M. DeFelice, I. Maayan, Y. Tanouchi, E. A. Ashley, and M. W. Covert, "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Computational Biology*, vol. 12, no. 11, 2016.

[62] S. U. Akram, J. Kannala, L. Eklund, and J. Heikkilä, "Cell proposal network for microscopy image analysis," in *International Conference on Image Processing*, pp. 3199–3203, IEEE, 2016.

[63] B. Dong, L. Shao, M. Da Costa, O. Bandmann, and A. F. Frangi, "Deep learning for automatic cell detection in wide-field microscopy zebrafish images," in *12th International Symposium on Biomedical Imaging*, pp. 772–776, IEEE, 2015.

[64] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, and Y. Fan, "Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks," *Neurocomputing*, vol. 229, pp. 88–99, 2017.

[65] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[66] C. Li, X. Wang, W. Liu, and L. J. Latecki, "Deepmitosis: Mitosis detection via deep detection, verification and segmentation networks," *Medical Image Analysis*, vol. 45, pp. 121–133, 2018.

[67] L. Chan, M. S. Hosseini, C. Rowsell, K. N. Plataniotis, and S. Damaskinos, "Histosegnet: Semantic segmentation of histological tissue type in whole slide images," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10662–10671, 2019.

[68] S. Takahama, Y. Kurose, Y. Mukuta, H. Abe, M. Fukayama, A. Yoshizawa, M. Kitagawa, and T. Harada, "Multi-stage pathological image classification using semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10702–10711, 2019.

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

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.

[71] T. de Bel, M. Hermsen, B. Smeets, L. Hilbrands, J. van der Laak, and G. Litjens, "Automatic segmentation of histopathological slides of renal tissue using deep learning," in *Medical Imaging 2018: Digital Pathology*, vol. 10581, p. 1058112, International Society for Optics and Photonics, 2018.

[72] K. R. Oskal, M. Risdal, E. A. Janssen, E. S. Undersrud, and T. O. Gulsrud, "A U-net based approach to epidermal tissue segmentation in whole slide histopathological images," *SN Applied Sciences*, vol. 1, no. 7, p. 672, 2019.

[73] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "A deep learning approach for semantic segmentation in histology tissue images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 176–184, Springer, 2016.

[74] A. Phillips, I. Teo, and J. Lang, "Segmentation of prognostic tissue structures in cutaneous melanoma using whole slide images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 0–0, 2019.

[75] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Medical Image Analysis*, vol. 36, pp. 135–146, 2017.

[76] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, I. Eric, and C. Chang, "Gland instance segmentation using deep multichannel neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 12, pp. 2901–2912, 2017.

[77] S. Graham, H. Chen, J. Gamper, Q. Dou, P.-A. Heng, D. Snead, Y. W. Tsang, and N. Rajpoot, "MILD-Net: minimal information loss dilated network for gland instance segmentation in colon histology images," *Medical Image Analysis*, vol. 52, pp. 199–211, 2019.

[78] A. BenTaieb, J. Kawahara, and G. Hamarneh, "Multi-loss convolutional networks for gland analysis in microscopy," in *13th International Symposium on Biomedical Imaging*, pp. 642–645, IEEE, 2016.

[79] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528, 2015.

[80] A. Arnab, S. Jayasumana, S. Zheng, and P. H. Torr, "Higher order conditional random fields in deep neural networks," in *European Conference on Computer Vision*, pp. 524–540, Springer, 2016.

[81] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, R. R. Gupta, Y. Gao, W. Chen, D. Foran, *et al.*, "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," *Pattern Recognition*, vol. 86, pp. 188–200, 2019.

[82] L. Sun, Z. Fan, X. Ding, Y. Huang, and J. Paisley, "Joint CS-MRI reconstruction and segmentation with a unified deep network," in *International Conference on Information Processing in Medical Imaging*, pp. 492–504, Springer, 2019.

[83] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," *arXiv preprint arXiv:1906.06876*, 2019.

[84] L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, "Unsupervised histopathology image synthesis," *arXiv preprint arXiv:1712.05021*, 2017.

[85] L. Bi, D. Feng, and J. Kim, "Dual-path adversarial learning for fully convolutional network (FCN)-based medical image segmentation," *The Visual Computer*, vol. 34, no. 6-8, pp. 1043–1052, 2018.

[86] F. Mahmood, D. Borders, R. Chen, G. N. McKay, K. J. Salimian, A. Baras, and N. J. Durr, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE Transactions on Medical Imaging*, 2019.

[87] B. Hu, Y. Tang, I. Eric, C. Chang, Y. Fan, M. Lai, and Y. Xu, "Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1316–1328, 2018.

[88] P. Luc, C. Couprie, S. Chintala, and J. Verbeek, "Semantic segmentation using adversarial networks," *arXiv preprint arXiv:1611.08408*, 2016.

[89] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, 2017.

[90] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[91] X. Zhu, X. Zhang, X.-Y. Zhang, Z. Xue, and L. Wang, "A novel framework for semantic segmentation with generative adversarial network," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 532–543, 2019.

[92] L. Mei, X. Guo, and C. Cheng, "Semantic segmentation of colon gland with conditional generative adversarial network," in *Proceedings of the 9th International Conference on Bioscience, Biochemistry and Bioinformatics*, pp. 12–16, 2019.

[93] A. B. Tosun, M. Kandemir, C. Sokmensuer, and C. Gunduz-Demir, "Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection," *Pattern Recognition*, vol. 42, no. 6, pp. 1104–1112, 2009.

[94] A. C. Ruifrok, D. A. Johnston, *et al.*, "Quantification of histochemical staining by color deconvolution," *Analytical and Quantitative Cytology and Histology*, vol. 23, no. 4, pp. 291–299, 2001.

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

[96] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[97] T. Gultekin, C. F. Koyuncu, C. Sokmensuer, and C. Gunduz-Demir, "Two-tier tissue decomposition for histopathological image representation and classification," *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 275–283, 2014.

[98] C. Schmid, "Constructing models for content-based image retrieval," in *Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II–II, IEEE, 2001.

[99] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol,

et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[100] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," in *International Conference Image Analysis and Recognition*, pp. 914–922, Springer, 2018.

[101] S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier, "Classification of breast cancer histology images using transfer learning," in *International Conference Image Analysis and Recognition*, pp. 812–819, Springer, 2018.

[102] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, *et al.*, "Detecting cancer metastases on gigapixel pathology images," *arXiv preprint arXiv:1703.02442*, 2017.

[103] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[104] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2015.

[105] A. Dadashzadeh and A. T. Targhi, "Multi-level contextual network for biomedical image segmentation," *arXiv preprint arXiv:1810.00327*, 2018.

[106] M. Z. Alom, T. Aspiras, T. M. Taha, V. K. Asari, T. Bowen, D. Billiter, and S. Arkell, "Advanced deep convolutional neural network approaches for digital pathology image analysis: A comprehensive evaluation with different use cases," *arXiv preprint arXiv:1904.09075*, 2019.

[107] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," in *Advances in Neural Information Processing Systems*, pp. 109–117, 2011.