Deep Multi-Task Learning for Histopathology Image Processing

A Dissertation

Presented in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

with a

Major in Computer Science

in the

College of Graduate Studies

University of Idaho

by

Haotian Wang

Major Professor: Min Xian, Ph.D.

Co- Major Professor: Constantinos Kolias, Ph.D.

Committee Members: Aleksandar Vakanski, Ph.D.; Tiankai Yao, Ph.D.

Department Administrator: Terence Soule, Ph.D.

May 2023

Abstract

Histopathological image analysis is challenging but essential for cancer detection, diagnosis, and prognosis prediction. Traditionally, pathologists examine the shapes and distributions of tissues under the microscope to identify the presence of carcinoma and quantify the level of malignancy. Because of the large number of tissue details, the whole process is time-consuming, low-throughput, and prone to human error. To overcome these challenges, computer-aided diagnosis (CAD) systems are leveraged to extract biologically relevant and human-interpretable features from histopathology images automatically. This automated process can significantly improve the performance of disease detection and enhance the reproducibility of results. However, the conventional single-task-based machine learning algorithms achieved poor generalizability in histopathology image analysis, particularly when only limited annotated histopathological data is available.

In this dissertation, I build a suite of deep multi-task learning (MTL) approaches to enhance the generalizability of machine learning models for the most challenging tasks in histopathology image processing. MTL uses multiple domain information-enriched tasks as inductive biases to improve the generalization of machine learning models. It enables machine learning models to learn shared representation from multiple tasks to exploit the commonalities and differences between them, which could greatly reduce models' dependency on large datasets. The main objective of this research is to utilize MTL neural networks to accurately extract objects of interest, such as glands and nuclei, in histopathology images and quantify the presence of tissue's morphology, topology, and geometry information. This information enables further analysis of digital biomarkers and is employed as the fundamental knowledge for cancer analysis.

First, I propose a novel MTL network with a bending loss regularizer to separate overlapped nuclei accurately. The newly proposed MTL architecture enhances the generalization by learning shared representation from three tasks: instance segmentation, nuclei distance map prediction, and overlapped nuclei distance map prediction. The proposed bending loss defines high penalties to concave contour points with large curvatures and applies small penalties to convex contour points with small curvatures. Minimizing the bending loss avoids generating contours that encompass multiple nuclei.

Second, I propose a novel topology-aware MTL network to accurately segment severely deformed and densely clustered glands. The proposed network has an MTL architecture and enhances the generalization of gland segmentation by learning shared representation from two tasks: instance segmentation and gland topology estimation. The proposed topology loss computes gland topology

using gland skeletons and markers. It drives the network to generate segmentation results that comply with the true gland topology.

Third, we propose a style-guided instance-adaptive normalization (SIAN) approach to synthesize realistic color distributions and textures for histopathology images from different organs. SIAN contains four phases, semantization, stylization, instantiation, and modulation. The first two phases synthesize image semantics and styles by using semantic maps and learned image style vectors. The instantiation module integrates geometrical and topological information and generates accurate nuclei boundaries. SIAN can generate realistic images from the layout to the histopathology images with target styles that align with different organs and cancer.

Lastly, I review MTL approaches for other image modalities and applications, and I demonstrate the effectiveness of MTL in Scanning Electron Imagining examination for metallic fuels.

Acknowledgments

First, I would like to express my heartfelt appreciation to the members of my Ph.D. committee. Dr. Min Xian has played a pivotal role in my academic and personal growth over the last five years. He has not only served as my major professor but also as an invaluable mentor who has supported and encouraged me every step of the way in my research and my life. I thank Dr. Constantinos Kolias for developing my research methodology and expertise, and Dr. Aleksandar Vakanski for generously sharing his time, expertise, and patience, and inspiring me to achieve excellence. I thank Dr. Tiankai Yao for his breadth and depth of experience to the committee. I would like to express my gratitude to the faculty and staff of the Department of Computer Sciences and the Idaho Falls center at the University of Idaho for providing me with a friendly, supportive, and intellectually stimulating graduate experience. Additionally, I extend my appreciation to my colleagues and project partners - Bryar Shareef, Fei Xu, Sujata Butte, Shoukun Sun, George Makrakis, and Ross Kunz - for their invaluable assistance.

Dedication

Most importantly, I am immensely grateful to my mother Wei Wu, and my father Xiangmeng Wang, for making innumerable sacrifices to provide me with a loving home and a quality education. Their unwavering guidance and support have been instrumental in helping me reach this point in my academic journey. I would also like to express my heartfelt appreciation to my loving wife, Ying Fang, for her constant support, love, and understanding throughout my studies and family life. Without the unwavering love of my parents and my wife, I wouldn't have been able to achieve what I have today.

Table of Contents

List of Tables

List of Figures

Statement of Contribution

This dissertation is submitted to the University of Idaho in partial fulfillment of the requirements for the degree of Doctor of Philosophy. I hereby declare that this work is my own and original work, and has not been previously submitted for obtaining an academic degree. All sources of information used have been appropriately acknowledged.

Chapter 2-4 of this dissertation are the portion of the published research. I would like to state the following contributions that any co-authors have made to the research published in this dissertation:

Chapter 2: Haotian Wang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing original draft, Visualization. Aleksandar Vakanski: Investigation, Writing – review & editing. Changfa Shi: Investigation. Min Xian: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision.

Chapter 3: Haotian Wang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing original draft, Visualization. Min Xian: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision. Aleksandar Vakanski: Investigation, Writing – review & editing.

Chapter 4: Haotian Wang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing original draft, Visualization. Min Xian: Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision. Aleksandar Vakanski: Investigation, Writing – review & editing. Bryar Shareef: Investigation.

Chapter 1: Introduction

1.1 Histopathology Image Processing

Cancer is a global public health problem and is one of the leading causes of death worldwide. According to the American Cancer Society, there were an estimated 1,918,030 new cancer cases and 609,360 cancer deaths in the United States in 2022 (Siegel et al., 2022). Histopathology offers the gold standard for cancer detection, diagnosis, and prognosis. It refers to the examination of tissues and cells at the microscopic level to detect diseases. Traditionally, pathologists visually examined histopathological samples, and the whole process was entirely manual and relied solely on the pathologist's expertise and experience. With the increasing number of cancer patients, tissue samples, and the complexity of diseases, pathologists face challenges in diagnosing cancer efficiently, accurately, and consistently.

Recently, machine learning (ML) has been increasingly adopted to assist with the diagnosis of cancer by automating routine tasks. ML uses algorithms to automatically learn patterns and insights from big data, instead of relying on pre-determined empirical equations. It offers a more accurate and efficient solution for analyzing and interpreting large volumes of histopathological image data, such as classifying, detecting, and segmenting specific structures or tissues within histopathology samples (Chen et al., 2016; Graham, Chen, et al., 2019; Graham, Vu, et al., 2019; Naylor et al., 2018; Oda et al., 2018; Qu et al., 2019; Vu et al., 2019; Xing et al., 2015; Yan et al., 2020; Zeng et al., 2019; Zhou et al., 2019). These methods show promising results and significant improvement in the speed of the diagnostic process and ensure consistent completion of tasks, which can reduce the workload of pathologists and allow them to focus on more complex cases that require their high-level professional expertise.

1.1.1 Major Challenges

Although encouraging progress has been achieved in ML-enhanced histopathology image analysis and processing, there still remain challenges for the computer-aided diagnosis system. The major challenges pertinent to hematoxylin and eosin (H&E)-stained histopathology image processing are summarized as follows.

1. **Accurate segmentation for overlapped and clustered tissues.** Overlapped and clustered tissues are one of the major challenges in histopathology image analysis (Graham, Vu, et al., 2019), where two or more different tissues overlap and cluster together in the image. This issue can be occurred by a variety of factors, such as tissue deformation (malignancy), pathological conditions, different tissue preparation procedures, and variations in imaging quality. Figure 1.1 illustrates an image
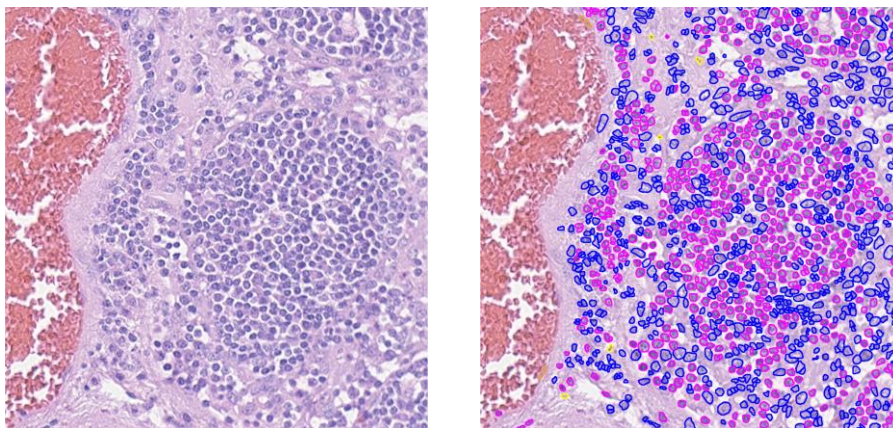
Figure 1.1 A sample H&E stained histopathology image patch (Graham, Vu, et al., 2019). Left shows the original image. Right shows the annotated nuclei in the image; different color refers to different types of nuclei.

patch that contains hundreds of nuclei that are overlapped and clustered in a small region, making it difficult to distinguish and classify each individual tissue. It presents a challenge not only for pathologists but also for machine learning algorithms to accurately segment and separate each individual nucleus. Separating overlapped tissues is important for accurate quantification and diagnosis. Overlapping tissues can cause errors in the quantification of various histopathological features, such as size, shape, and texture. If two or more objects are detected as a single object, their individual characteristics may be lost, leading to inaccurate measurements. For example, in cancer diagnosis, accurate counting and characterization of the nuclei are critical for determining the grade and stage of cancer.

2. **Accurate segmentation for malignant tissues.** Malignant tissues are characterized by abnormal and uncontrolled cell growth that can invade nearby tissues and spread to other parts of the body (Epstein et al., 2005). These tissues are cancerous or have the potential to become cancerous, and they often appear irregular in shape and size. The irregularity of malignant tissues poses challenges to accurately distinguishing and segmenting them. In histopathology images, healthy glands typically appear in round or oval shapes and are uniform in appearance in terms of size and shape. In contrast, malignant tissues tend to grow oversize, and invade adjacent tissues, causing severe deformation of their shapes, and cluster densely in histopathology images (Figure 1.2). The accurate detection and treatment of malignancy is crucial for improving patient outcomes and survival rates. Early diagnosis and prompt treatment are important for managing malignant tumors and preventing them spread to other parts of the body. However, accurately segmenting malignant
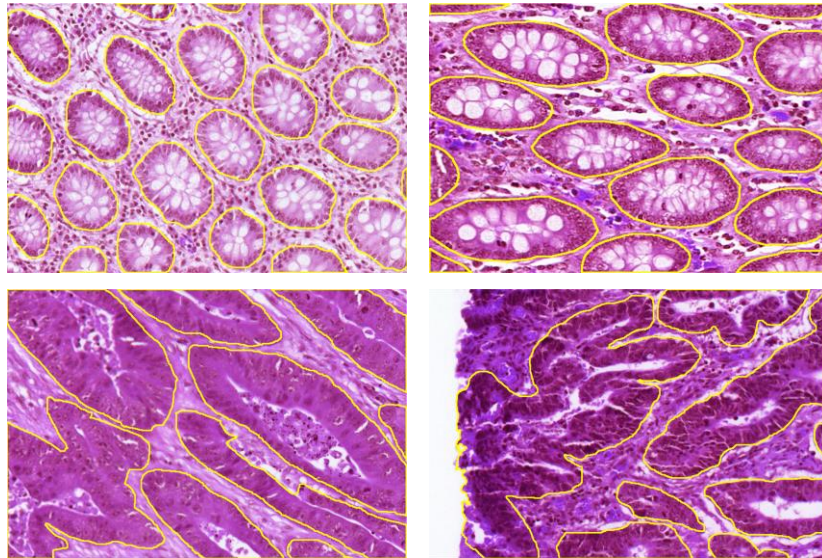
Figure 1.2 The H&E stained histopathology images show healthy and malignant glands with yellow contours (Sirinukunwattana et al., 2015). The first row displays healthy glands, the second row shows malignant glands. The malignant glands are appeared in deformed shape and in close proximity to their neighboring glands.

tissues can be challenging due to their large variation in irregularities such as shape, size, alignment, and nuclear morphology.

3. **Color variation among various cancer and organs.** Cancers are diagnosed by the first developed organ (primary site) or the tissue type where cancer originated (histological type). In histopathology images, different types of cancer can have large variations in color and texture (Figure 1.3). Color variations can arise from different structure types, staining techniques, imaging settings, and processing methods (Veta et al., 2014). For example, the choice of staining techniques and dyes used, different imaging settings, and processing techniques may vary among different tissue types and cancers, ultimately leading to variations in the color of histopathology images. This color variation can impact the accuracy of detection and segmentation algorithms, which in turn affects the effectiveness of diagnosis and treatment.
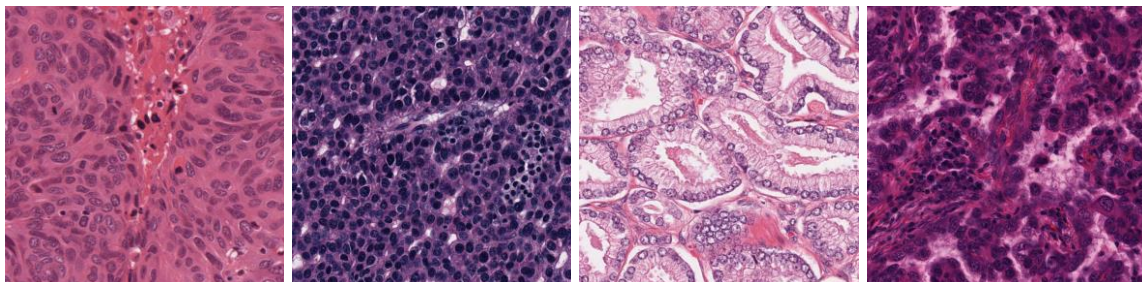
Figure 1.3 Color variations among H&E stained histopathology images (Kumar et al., 2017).

4. **Limited annotated samples.** Supervised learning algorithms have demonstrated the b est overall performance in histopathology image processing, but they require a substantial amount of annotated data for effective model training and convergence. Current large-scale image datasets, such as ImageNet (Deng et al., 2009), and Microsoft COCO (Lin et al., 2014), contain hundreds of thousands to millions of samples. However, existing Hematoxylin and eosin (H&E) stained histopathological image datasets are comparatively limited, and only tens to hundreds of image patches were used in Graham, Chen, et al. (2019); Graham, Vu, et al. (2019); Kumar et al. (2017); Naylor et al. (2018); Vu et al. (2019); Kumar et al. (2019); Sirinukunwattana et al. (2015). Consequently, the current heavily over-parameterized learning models may overfit the limited training data and lack generalizability on new and unseen data. Figure 1.1 shows a sample of H&E stained histopathology image patch from the CoNSeP dataset (Graham, Vu, et al., 2019), which contains hundreds of nuclei in a 1000×1000 pixel patch. Thus, annotating large histopathology image datasets with pixel-accurate annotations can be extremely costly, especially for images that contain substantial amounts of tissues.

These challenges present researchers with opportunities to develop innovative solutions. In the dissertation, we focus on these four major challenges, which are addressed in different chapters of the study. To tackle these challenges, multi-task learning is employed as a primary method to advance our understanding of histopathology image processing and contribute to the development of more accurate and efficient methods for histopathology image analysis.

1.2 Multi-task Learning

Multi-task learning (MTL) is a machine learning approach that aims to train a model to learn multiple independent tasks simultaneously. MTL comprises shared feature extractor layers that extract shared features from input data and multiple tasks that produce predictions for each task-specific output. This design exploits the similarities and differences among tasks, enabling the transfer of knowledge between them, which improves model generalization, strengthens latent representations, and enables domain adaptation (Caruana, 1998). The learning procedure mimics how humans excel at recognizing similarities between new problems and prior experiences.

There are two main categories of existing MTL methods: hard-parameter sharing and soft-parameter sharing (Crawshaw, 2020; Ruder, 2017). Hard-parameter sharing is the most commonly used strategy to perform MTL in deep neural networks. In this approach, shared hidden layers between tasks are jointly learned in a representation learning subnetwork, and multiple subnetworks are used to predict each task-specific output, as illustrated in Figure 1.4 (a). In soft parameter sharing, each task

(a) Hard-parameter Sharing MTL

(b) Soft-parameter Sharing MTL

(c) Multiple Single Task Learning

(d) Multi-label Learning
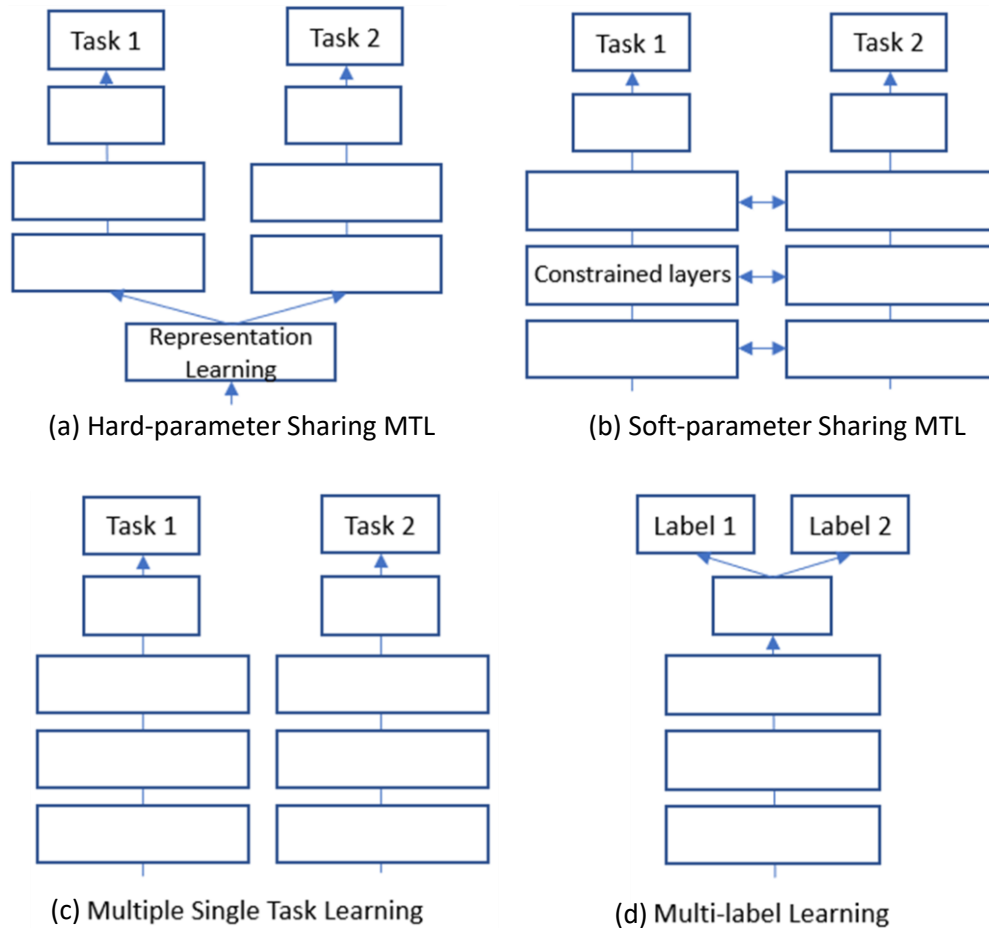
Figure 1.4 Deep neural network architectures can predict multiple tasks.

has its own model with its own parameters, where the associated parameters are highly constrained, and they could learn similar knowledge for each task, as evidenced in Figure 1.4 (b). However, soft-parameter sharing can limit the learning of dissimilar tasks due to its constrained layers and introduce more computational costs when the number of tasks is large. Therefore, hard-parameter sharing MTL is the most prevalent design in recent studies.

It is crucial to distinguish MTL from other deep learning architectures that can predict multiple tasks, such as multiple single-task learning (MSTL) and Multi-label learning (MLL). MSTL uses multiple single-task learning models, one for each task, to predict multiple tasks simultaneously, as illustrated in Figure 1.4 (c). In the inference, these models are combined to predict multiple tasks. MSTL suffers from a lack of shared knowledge across multiple tasks and requires more computational resources as the number of tasks increases. MLL involves each data point being associated with multiple labels. If each label is treated as a separate task, it can predict multiple tasks using MLL, as depicted in Figure 1.4 (d). However, learning multiple labels using one architecture could limit performance if different labels have relatively unrelated tasks, such as predicting classification labels

and segmentation labels simultaneously. In this dissertation, we focus on **hard-parameter sharing MTL**, which is the most prevalent and effective design in comparison to other MTL architectures such as MSTL, MLL, and soft-parameter sharing MTL. For simplicity, the term "MTL" refers to the hard-parameter sharing MTL throughout the rest of this dissertation.

In formulating MTL for image recognition, we define $\boldsymbol{x}$ be the input data, and $\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n}$ be the output tasks. We aim to learn an MTL function f that can simultaneously map the input data to multiple outputs by minimizing the loss/error between $f(\boldsymbol{x})$ and $\boldsymbol{y_i}$. The MTL learning objective is defined as follows:

$$\mathcal{L}(f(\boldsymbol{x}, \boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n})) = \sum_{i=1}^{n} \lambda_i \, \mathcal{L}_i(f(\boldsymbol{x}), \boldsymbol{y_i}) + \ell(f)$$

$$(1.1)$$

where $\mathcal{L}_i$ is a pre-defined task-dependent loss function for task i, such as cross-entropy for a discrete target or mean squared error for a continuous target, λ_i is a weighting factor for each task, $\ell(f)$ is the regularization term to prevent overfitting and improve the generalization performance of the model. The first term in the loss function is the overall summation for each task-specific loss, controlled by weighting factor λ_i. The second term, $\ell(f)$, encourages the model to learn a shared representation across all tasks, which helps to improve generalization performance. During training, the network minimizes the loss function for all tasks simultaneously, encouraging the network to learn task-specific representations that are also informative for other tasks. The shared feature representations enable the model to generalize better to new tasks, while the task-specific output layers allow the model to make multiple task-specific predictions. By jointly learning multiple tasks, multi-task learning can improve data efficiency, improve model generalizability, reduce the risk of overfitting, and introduce faster convergence compared to learning each task independently.

One important advantage of MTL is to alleviate the data sparsity problem, as it improves data efficiency and enhances model generalizability. Each task has a limited number of annotated data could be insufficient to train an accurate model. MTL addresses this problem by aggregating the annotated data from all the tasks to train a more accurate model. This can help reuse existing knowledge and reduce the cost of manual annotations for model training. Furthermore, MTL benefits from diverse data from different learning tasks. This leads to learning more robust and universal representations from different tasks, resulting in more accurate and effective models. Recent studies have demonstrated the effective accuracy, efficiency, and generalizability of MTL in various fields, including natural language processing (Worsham & Kalita, 2020), computer vision (Kendall et al., 2018), and biomedical image analysis (Chen et al., 2016; Graham, Vu, et al., 2019).
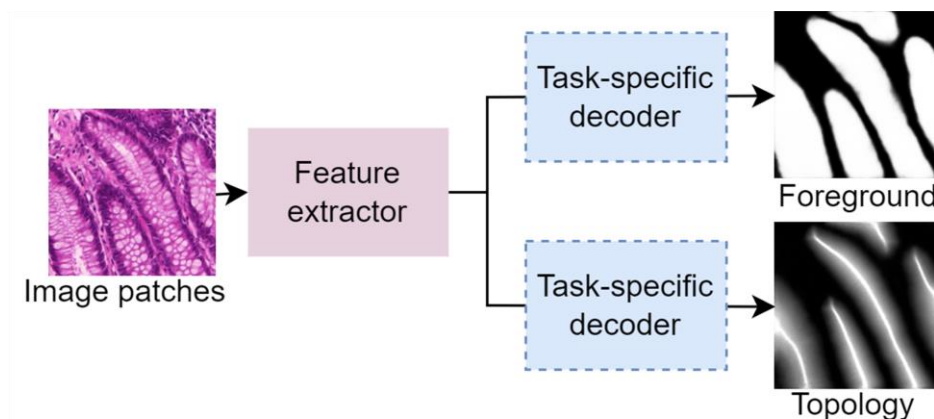
Figure 1.5 A sample multi-task learning network in histopathology images. The MTL takes histopathology image inputs into a shared feature extractor and learns multiple tasks simultaneously with task-specific outputs, such as foreground and topology.

In histopathological image processing, MTL is particularly advantageous because it can address multiple tasks simultaneously to enhance model generalizability, such as object segmentation, object classification, and object morphology and topology. This design is important to address various histopathological challenges, such as malignant tissues and overlapped and clustered tissues, by integrating sufficient topological, morphological, and geometrical features to the common features, such as object foreground, object attributes, allowing the model to tackle the challenges from different prospects. Furthermore, in the case of limited annotated histopathological data, MTL can improve data efficiency from the data provided by other tasks and enable more effective transfer learning between tasks. Additionally, MTL helps to reduce the risk of overfitting on any specific task and facilitates faster convergence on the network since the network can utilize the diverse data attributes to learn the shared feature representations. Figure 1.5 illustrates an MTL network that takes advantage of foreground segmentation, and topology estimation simultaneously for accurate gland segmentation.

To design an effective MTL, several factors need consideration while creating MTL and its shared architecture, such as the portion of the parameters to share between tasks, the architecture design of parameterization, and the combination of task-specific and shared modules (Vandenhende et al., 2021). Existing proposed MTL architectures involve a trade-off in the degree of information sharing between tasks. Over-sharing can lead to negative transfer, resulting in a worse performance of the joint multi-task model than multiple individual models for each task. Conversely, limiting sharing may hinder the model from leveraging information effectively between tasks. Thus, the optimal performance of MTL architecture strikes an effective balance in sharing information between tasks.

1.3 Contribution

The main objective of this research is to utilize multi-task learning neural networks to accurately extract structures within tissue, such as glands and nuclei, and analyze their morphology for accurate cancer detection. The study primarily employs machine learning algorithms and key concepts from multi-task learning theory. This thesis is organized into chapters that address various cancer diseases and challenges related to histopathological diagnosis.

In Chapter 2, we introduce the novel deep multi-task networks with bending loss regularizer for accurate nuclei segmentation and overlapped nuclei segmentation:

In Chapter 3, we propose a novel deep multi-task network to learn the gland instance and gland topological and morphological information simultaneously for accurate gland segmentation and deformed gland segmentation.

In Chapter 4, we propose a deep generator network that can accurately generate realistic histopathology images and align images to different disease styles to increase the histopathology data and improve the supervised training performance.

Chapter 5 discusses the application of the proposed network in other image domains and image modalities, such as scanning electron microscopy images.

Overall, our studies present convincing visual representations and promising statistical improvements in many histopathology image applications including nuclei segmentation, gland segmentation, histopathology image generation, and realistic manipulation of histopathological textures, and the images in other scientific domains.

Chapter 2: Bending Loss Regularized Multi-Task Learning Network for Histopathology Nuclei Segmentation

Wang, H., Xian, M., & Vakanski, A. (2020). Bending loss regularized network for nuclei segmentation in histopathology images. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI),

Wang, H., Vakanski, A., Shi, C., & Xian, M. (2021). Bend-Net: Bending Loss Regularized Multitask Learning Network for Nuclei Segmentation in Histopathology Images. arXiv preprint arXiv:2109.15283.

2.1 Introduction

Histopathology nuclei analysis provides direct and reliable evidence for cancer detection. Nuclei are the most distinguishable structures in H&E-stained histopathology images, and accurate segmentation of nuclei is essential for various quantitative analyses such as movement tracking, morphological changes, and nuclei counting. Traditionally, pathologists inspect nuclei morphology and distributions under microscopes to diagnose carcinoma and assess malignancy level. However, the large number of nuclei makes the whole process inefficient and prone to human error. Therefore, automated and accurate nuclei segmentation is essential for efficient and accurate cancer detection and diagnosis in clinical practice.

Traditional approaches (Ali & Madabhushi, 2012; Cheng & Rajapakse, 2008; Yang et al., 2006) utilized thresholding and watershed algorithms to segment nuclei, but these approaches have limited robustness in handling images with various nucleus types, fat tissue, and staining procedures. In recent years, deep learning-based approaches have been successfully applied to biomedical image processing tasks (Badrinarayanan et al., 2017; Long et al., 2015; Ronneberger et al., 2015), and nuclei segmentation (Chen et al., 2016; Graham, Chen, et al., 2019; Graham, Vu, et al., 2019; Naylor et al., 2018; Oda et al., 2018; Qu et al., 2019; Vu et al., 2019; Xing et al., 2015; Yan et al., 2020; Zeng et al., 2019; Zhou et al., 2019). For instance, Xing et al. (2015) proposed a convolution neural network (CNN) that produces probability maps, and improved robustness with postprocessing techniques such as distance transformation, H-minima thresholding, and region growing. Kumar et al. (2017) presented a three-class CNN that computes the label for each pixel as an instance, boundary, or background for nuclei segmentation. Naylor et al. (2018) proposed a deep regression network that utilized the nuclei distance representations to separate overlapped nuclei for accurate nuclei segmentation. Although these

(a) Image patch   (b) Ground truth   (c) U-Net   (d) SegNet

(e) DCAN   (f) DIST   (g) HoVer-Net   (h) Bend-Net
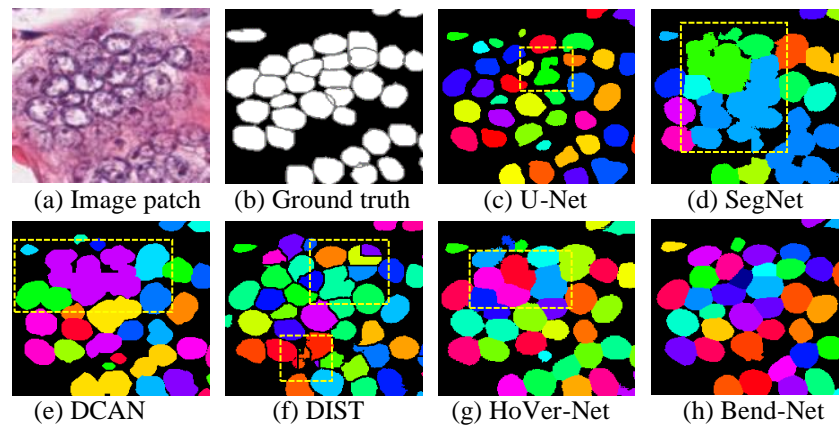
Figure 2.1 Examples of state-of-the-art approaches in segmenting overlapped nuclei.

methods have demonstrated better results compared to traditional approaches, however, it is still challenging to segment nuclei accurately due to the presence of a large amount of overlapped nuclei.

Overlapped nuclei segmentation is challenging due to the lack of clear boundaries among nuclei, similar background textures, and large size and morphology variations. Recently, deep learning-based approaches have been proposed to address this challenge using three main strategies. The first strategy utilized the neural network to split the overlapped nuclei by generating both nuclei instances and boundaries. For example, Kumar et al. (2017) proposed a three-classes CNN, including instances, boundaries, and background to segment the overlapped nuclei. Chen et al. (2016) proposed a multi-task learning framework that outputs instance map and boundary map in separate branches. Vu et al. (2019) constructed a multiscale deep residual network with instances and boundary classes to segment nuclei. The second strategy integrated features from overlapped nuclei to improve overall segmentation performance. Zhou et al. (2019) proposed the CIA-Net that utilized spatial and texture dependencies between nuclei and contours to improve the robustness of nuclei segmentation. Alemi Koohbanani et al. (2019) proposed a SpaNet that captures spatial features in a multiscale neural network. Graham and Rajpoot (2018) proposed a new weighted cross-entropy loss that was sensitive to the Hematoxylin stain. Qu et al. (2019) constructed a loss to learn spatial features for improving localization accuracy. The third strategy utilized the deep regression networks and Watershed algorithm to segment the overlapped nuclei. Naylor et al. (2018) constructed a regression network that generated markers for the Watershed algorithm. Graham, Vu, et al. (2019) proposed the HoVer-Net architecture to output the instance map and horizontal and vertical nuclei distance maps for obtaining the markers of the Watershed algorithm. According to the reported results, these approaches achieved better overall performance than traditional methods, but their ability to separate overlapped nuclei is still limited, as shown in Figure 2.1.

Figure 2.2 Two contours. (a) An ideal nucleus contour; and (b) a contour contains two nuclei. Red rectangles highlight the touching points on the contour.

To address the challenges above, we proposed a novel bending loss regularized deep multi-task network for nuclei segmentation. First, our proposed multi-task network comprises three decoder branches: 1) instance segmentation branch, 2) boundary-distance branch for all nuclei, and 3) boundary-distance branch for overlapped nuclei, which is designed to identify overlapped nuclei. The proposed multi-task network aims to learn a shared representation from three nuclei representations for accurate nuclei segmentation. Second, we propose the bending energy-based regularizer to penalize large curvatures of nuclei contours. In histopathology images, the curvatures of nucleus contour points change smoothly; but if one contour contains two or multiple overlapped or touching nuclei, their touching points on the contour will have sharp curvature changes (Figure 2.2). Inspired by this observation, we develop the bending loss to generate large penalties for contour points with large curvatures. Additionally, we propose two new metrics to evaluate overlapped nuclei segmentation. Previous approaches evaluate overlapped nuclei segmentation using metrics for overall segmentation performance, which hides the real performance of the overlapped nuclei segmentation. Finally, we validate the proposed approach on the CoNSeP and MoNuSegv1 datasets using seven quantitative metrics: Aggregate Jaccard Index, Dice, Segmentation Quality, Recognition Quality, Panoptic Quality, AJIO, and ACCO. Extensive experiments demonstrate that the proposed Bend-Net outperforms eight state-of-the-art approaches.

Compared to the closest work, HoVer-Net (Graham, Vu, et al., 2019), both the proposed approach and the HoVer-Net follow the multi-task learning architecture and use ResNet-50 as building blocks. There are two major differences between the two approaches: firstly, our proposed method includes a new decoder branch that specifically focuses on the segmentation of overlapped nuclei; secondly, we introduce the bending loss, which penalizes large curvatures of nuclei contours, to improve the accuracy of segmentation in cases where multiple nuclei are in close proximity or overlapping.

2.2 Bend-Net Method

The proposed method, namely Bend-Net, consists of two key components: the bending loss and multi-task learning architecture. Firstly, we propose a bending energy-based regularizer for penalizing touching nuclei points. Secondly, we propose a multi-task learning network with three decoder branches that focus on overlapped nuclei contours. The final loss function consists of the regular segmentation loss (Graham, Vu, et al., 2019), overlapped nuclei loss, and bending loss.

2.2.1 Bending Loss

Bending energy has been widely applied in measuring the shapes of biological structures, e.g., blood cells (Canham, 1970), cardiac (Duncan et al., 1991), vesicle membranes (Du et al., 2006), and blood vessels (Stuhmer et al., 2013). Young et al. (1974) used chain-code representations to model bending energy. Verbeek and Van Vliet (1993) used the derivative-of-Gaussian filter to model bending energy in the gray-scale image for motion tracking. Bergou et al. (2008) modeled the discrete curvature and bending loss both in kinematic and dynamic treatment to solve the smoothness problem.

For 2D digital images, a contour is composed of discrete pixels, and the curvature of a contour point is computed by using the vectors created by neighboring points on the contour. For histopathology images, a nucleus usually has a smooth contour, and the points on the contour have small curvature changes; the points on the contour with large curvature have a high probability to be the touching points of two/multiple nuclei (Figure 2.2). To split the touching nuclei, we define the bending loss that gives high penalties to the contour points with large curvatures and small penalties to points with small curvatures. The proposed total loss is given by

$$\mathcal{L} = \mathcal{L}_0 + \alpha \cdot \mathcal{L}_{be} \tag{2.1}$$

where L_0 refers to conventional segmentation loss (Section 2.3); $\mathcal{L}_{be}$ denotes the proposed bending loss; and the parameter α controls the contribution of the bending loss. Let $C = \{c_i\}_{i=1}^m$ be the set of contour points of nuclei in an image, and $\mathcal{L}_{be}$ is defined by

$$\mathcal{L}_{be}(C) = \frac{1}{m}\sum_{i=1}^m BE(i) \tag{2.2}$$

where $BE(i)$ is the discrete bending energy at the point c_i,

$$BE(i) = \frac{\kappa(i)^2 \left(\left(1 - \delta(c_i)\right) + \delta(c_i) \cdot \mu \right)}{|v(i, i+1)| + |v(i-1, i)|}$$

$$\tag{2.3}$$

| 0/0 | 0.28/5.69 | 1.41/28.28 | 2.00/40.00 | 9.66/193.14 |
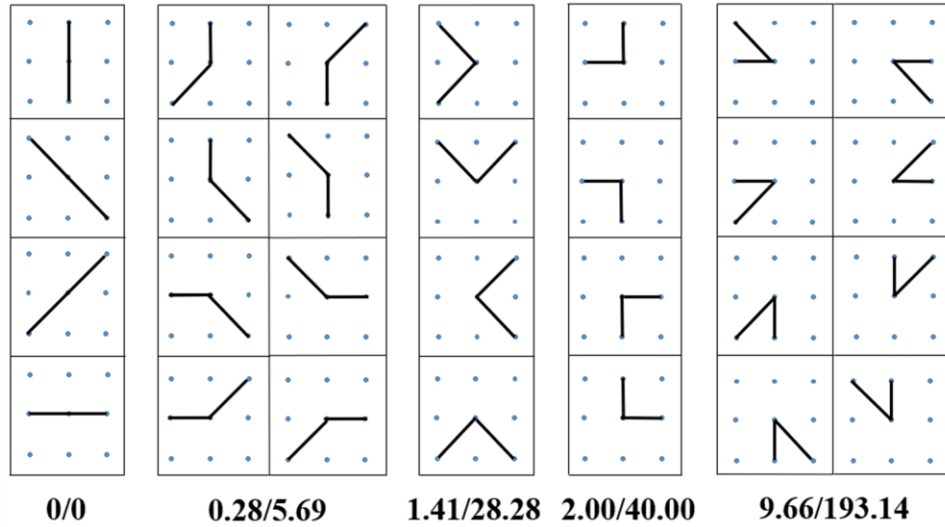
Figure 2.3 Discrete bending losses for different curve patterns. In the value pairs 'A/B', 'A' represents the convex bending loss for the center point, and 'B' denotes the concave bending loss. The value rounds to two decimal places.

$$\kappa(i) = \frac{2|v(i-1,i) \times v(i,i+1)|}{|v(i-1,i)||v(i-1,i)| + v(i-1,i) \cdot v(i,i+1)}$$

(2.4)

in Eq. (2.3), $\delta(c_i)$ is 1 if c_i is a concave contour point, and 0 if c_i is a convex point; $\kappa(i)$ is the curvature at c_i. For three consecutive pixels on a nucleus boundary with coordinates x_{i-1}, x_i and x_{i+1}, $v(i-1, i)$ is the edge vector from point $i-1$ to i, such that $v(i-1, i) = x_i - x_{i-1}$; and $v(i, i+1)$ is the edge vector from i to $i+1$, such that $v(i, i+1) = x_{i+1} - x_i$. Operator $|\cdot|$ calculates the length of a vector. μ defines the weight for concave contour points.

The 8-neighborhood system is applied to search neighbors for contour points. Ideally, a contour point only has two neighboring points, and their coordinates are used to calculate the edge vectors in Eqs. (2.3) and (2.4). As shown in Figure 2.3, a point with eight neighbors has 28 combinations of possible curve patterns. All curve patterns are divided into five groups; in each group, the concave points and the convex points have different discrete bending loss values. In the first group, the four patterns construct straight-line segments, and their bending losses are all 0s. The second group shows patterns with a $3\pi/4$ angle between ed ge vectors, and their bending losses are relatively small. In the last group, the eight patterns have large curvatures, and their bending losses are the largest in all patterns. The third and fourth groups illustrate patterns with the same angles between edge vectors, but they have different bending losses due to the different vector lengths.

To determine the concave and convex points, the mid-point of two extended neighboring points is calculated. If the mid-point is out of the predicted nucleus, we define it as a concave point; otherwise,
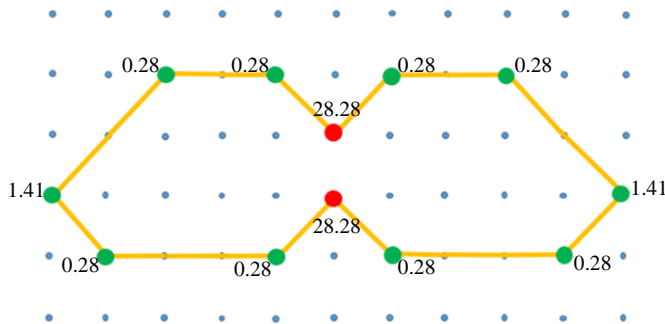
Figure 2.4 A contour with both concave and convex bending energy points. Red dots highlight the concave points, and green dots highlight the convex points.

(a) Nuclei contours   (b) Poorly-segmented   (c) Well-segmented
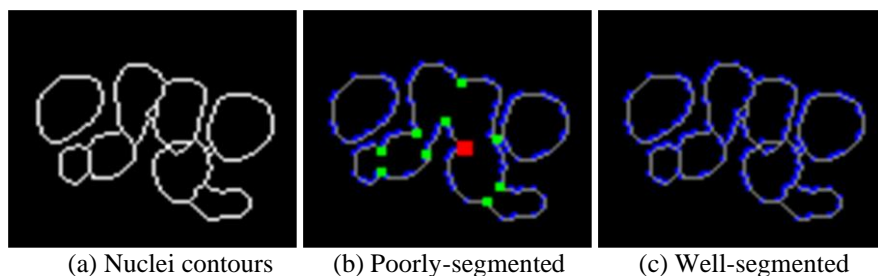
Figure 2.5 Different bending losses of different segmentation results. (a) Ground truth of eight nuclei contours; (b) bending losses of contour points of poorly-segmented nuclei; and (c) bending losses of well-segmented nuclei. Red: BE = 193.14, green: BE=28.28 and BE=40.0, blue: BE≤9.66, and grey: BE=0.

the point is a convex point. The concave points are more likely to be overlapped contour points, and the convex points are usually regular/normal points. Eq. (2.3) gives a larger penalty to concave points. The previous approach (Wang et al., 2020) calculated bending loss using curvature directly. Points with the same curvatures could be convex or concave, convex points are more likely regular contour points, and concave points are likely to be overlapped contour points. The previous approach cannot distinguish convex and concave contour points and tends to over-segment nuclei.

A sample of overlapped nucleus contour is shown in Figure 2.4. The red dots highlight the concave points and the green dots highlight the convex points. The concave points' bending loss values are 28.28. The mid-points of green dots are inside of the predicted nucleus, and they are convex points; their bending loss values are less than 1.41. In Figure 2.4, the concave points with a bending loss of 28.28 and the convex points with a bending loss of 1.41 have the same curve pattern; however, the concave points produce 20 times as much loss as the convex points.

The proposed bending loss is rotation invariant since all patterns with the same angle between two edge vectors have the same bending loss. In practice, if two nuclei contours share some contour segments, one contour point may have more than two neighbors. In this scenario, we calculate the
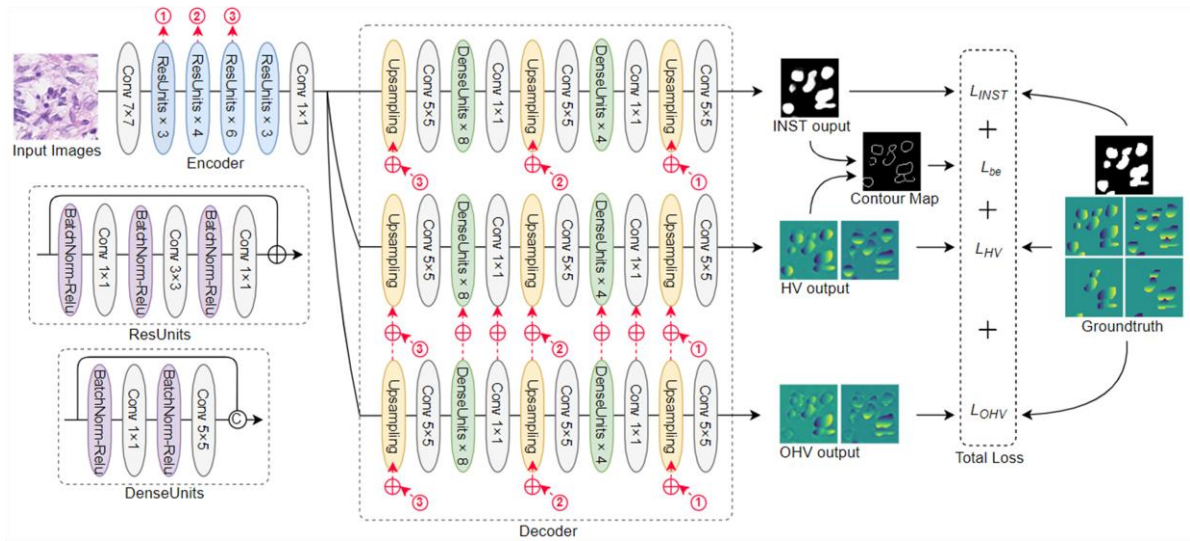
Figure 2.6 Overview of the proposed Bend-Net. ⊕ denotes the summation; © denotes the concatenation; red arrow represents the skip connections; number with red circle denotes the connected position of skip-connections.

bending loss for all possible combinations and choose the smallest loss as the discrete bending loss for the point.

As shown in Figure 2.5, for poorly-segmented nuclei contours, all these touching contour points have relatively high (red and green points) bending losses. If the touching nuclei are well separated (Figure 2.5(c)), and the bending loss of all contour points are less than 9.66.

2.2.2 Multi-task Learning Network

The proposed multi-task learning architecture is shown in Figure 2.6. The network follows an encoder-decoder design, and has three decoder branches. The encoder employs ResNet-50 (He et al., 2016) as a feature extractor. In the first convolutional layer, 64 7×7 kernels with a stride of 1 are applied, but the following max-pooling layer is removed to preserve more information. The network has three decoders/tasks. The first task predicts the nuclei instance map (INST); the second produces each nucleus's horizontal and vertical boundary-distance map (HV); and the third outputs the overlapped nuclei's horizontal and vertical boundary-distance map (OHV). All decoders in the three branches have the same sub-architectures and dense units. The OHV and HV branches share weights through skip connections.

The weight-sharing among decoders is designed to use features learned from similar tasks. In traditional multi-task learning networks, different branches typically addressed different tasks. However, in the proposed network, both the HV and OHV branches share some comment results; one for all nuclei, and the other for overlapped nuclei. To take advantage of the features from two similar tasks, we design skip connections among two branches to share weights. Specifically, the network first

learns the distance maps of overlapped nuclei and aggregates them through skip connections to distance maps in the HV branch.

2.2.3 Loss Function

As shown in Figure 2.6, the loss function of the proposed network has four terms: the losses from three different decoders and the proposed bending loss. Let $\mathcal{L}_{INST}$ denote the loss of the binary instance map; $\mathcal{L}_{HV}$ be the loss of the horizontal and vertical distance maps from the HV branch; and $\mathcal{L}_{OHV}$ denote the loss of the horizontal and vertical distance maps from the OHV branch; $\mathcal{L}_{be}$ is the bending loss. The proposed loss function also can split into the segmentation loss ($\mathcal{L}_0$) and the bending loss regularizer (Eq. 2.1). The total loss is given by:

$$\mathcal{L} = \underbrace{\mathcal{L}_{INST} + \mathcal{L}_{HV} + \mathcal{L}_{OHV}}_{\mathcal{L}_0} + \alpha \cdot \mathcal{L}_{be} \tag{2.5}$$

where α is the weight of the bending loss in all loss functions. We follow the design in (Graham et al., 2019) to set the losses of the three branches to have equal contributions to the total loss.

Loss of the INST branch. To segment the nuclei instance, we calculate the binary classification for each image pixel. I and I^* are the predicted instance map and the ground truth instance map for all nuclei. The loss ($\mathcal{L}_{INST}$) is a summation of the cross-entropy loss ($\mathcal{L}_{CE}$) and Dice loss ($\mathcal{L}_{Dice}$). They are given by

$$\mathcal{L}_{INST}(I, I^*) = \mathcal{L}_{CE}(I, I^*) + \mathcal{L}_{Dice}(I, I^*) \tag{2.6}$$

$$\mathcal{L}_{CE}(I, I^*) = -\frac{1}{n}\sum_i^n I_i^* \log(I_i) \tag{2.7}$$

$$\mathcal{L}_{Dice}(I, I^*) = 1 - \frac{2 \times \sum_i^n I_i I_i^*}{\sum_i^n I_i + \sum_i^n I_i^*} \tag{2.8}$$

where I_i is the class prediction at point i, and n denotes the number of pixels in an image patch. The INST branch separates the nuclei instance from the background.

Loss of the HV branch. The loss function is to compare the predicted distance maps (D) with the ground truth distance maps (D^*) for all nuclei. We employed the distance loss function in (Graham et al., 2019). The distance loss function is defined by

$$\mathcal{L}_{dist}(D, D^*) = \mathcal{L}_{Mse}(D, D^*) + 2 \cdot \mathcal{L}_{Msge}(D, D^*) \tag{2.9}$$

$$\mathcal{L}_{Mse}(D, D^*) = \frac{1}{n} \sum_i^n d_i^2 \tag{2.10}$$

$$\mathcal{L}_{Msge}(D, D^*) = \frac{1}{n} \sum_{i}^{n} \left(\nabla_{d_i}\right)^2 \tag{2.11}$$

where $\mathcal{L}_{Mse}$ is the mean square error loss and $\mathcal{L}_{Msge}$ is the mean square gradient error loss; d is $D - D^*$, and ∇ denotes the gradient calculation.

Loss of the OHV branch. $\mathcal{L}_{OHV}$ is also defined using mean square error and the mean square gradient error (Eq. 2.9). But $\mathcal{L}_{OHV}$ is calculated using the predicted distance maps (D) and the ground truth distance maps of overlapped nuclei.

2.3 Experimental Results and Discussion

In this section, we first discuss the datasets, evaluation metrics, and experimental settings. Then, we show the effectiveness of each component of the Bend-Net. Then, we compared it to state-of-the-art methods using overall segmentation performance, and overlapped nuclei segmentation.

2.3.1 Datasets and Evaluation Metrics

We validate the proposed method using two histopathology nuclei datasets: CoNSeP (Graham, Vu, et al., 2019) and MoNuSegv1 (Kumar et al., 2017). CoNSeP is provided by the University of Warwick and has 41 H&E-stained images from 16 colorectal adenocarcinomas (CRA) WSIs collected using the Omnyx VL120 scanner. Six types of nuclei, normal epithelial, tumor epithelial, inflammatory, necrotic, muscle, and fibroblast exist in the dataset. The dataset contains 24,319 manually annotated nuclei (13,256 overlapped). The image size is 1000×1000 with the magnification at $40\times$. In the experiment, 27 images are utilized for training and validation, and 14 images for testing. The training and validation sets have 15,582 nuclei, and the test set has 8,791 nuclei.

MoNuSegv1 contains 30 images from TCGA (The Cancer Genomic Atlas) dataset. The original size of the images is 1000×1000, and there are more than 21,000 manually annotated nuclei from the breast, liver, kidney, prostate, bladder, colon, and stomach. The magnification is at $40\times$. In experiments, 16 images (4 breasts, 4 livers, 4 kidneys, 4 prostates) are used for training and validation, and 14 images for testing. The training and validation sets contain over 13,000 nuclei (4,431 overlapped), and the test set has 6,000 nuclei (2,436 overlapped). The author recently extended the dataset and published in (Kumar et al., 2020); however, it was not adopted in this study because it contains much less overlapped nuclei in their new test set compared with the previous test set (Kumar et al., 2017).

We employed five quantitative metrics to evaluate the overall performance of nuclei segmentation approaches: Aggregate Jaccard Index (AJI) (Kumar et al., 2017), Dice coefficient (Dice, 1945), Recognition Quality (RQ) (Kirillov et al., 2019), Segmentation Quality (SQ) (Kirillov et al., 2019), and Panoptic Quality (PQ) (Kirillov et al., 2019). We propose two new metrics to evaluate the overlapped

nuclei segmentation: Aggregated Jaccard Index of overlapped nuclei and accuracy for overlapped nuclei.

Let $G = \{G_i\}_{i=1}^{N}$ be the nuclei ground truth of an image, N denotes the total amount of segments in G; and let $S = \{S_k\}_{k=1}^{M}$ be the predicted segments of the corresponding image, M denotes the total amount of segments in S. AJI is an aggregate version of the Jaccard Index and is defined by

$$\text{AJI} = \frac{\sum_{i=1}^{N} G_i \cap S_j}{\sum_{i=1}^{N} G_i \cup S_j + \sum_{S_k \epsilon U} S_k}$$

where S_j is the matched predicted segments that produce the largest Jaccard Index value with G_i; and U denotes the set of unmatched predicted segments, where the total amount of U is $(M - N)$.

Dice coefficient (DICE) is utilized to evaluate overall segmentation performance, the DICE is given by

$$DICE = \frac{2|G \cap S|}{(|G| + |S|)}$$

where operator $|\cdot|$ denotes the cardinalities of the segments.

PQ is used to estimate both detection and segmentation results. RQ is the familiar F1-score, and SQ is known as the average Jaccard Index of matched pairs. *RQ, SQ, PQ* are defined as

$$RQ = \frac{TP}{TP + \frac{1}{2}FP + \frac{1}{2}FN}$$

$$SQ = \frac{\sum_{(p,g)\in TP} IoU(p,g)}{TP}$$

$$PQ = RQ \times SQ$$

where p refers to prediction, g refers to the ground truth. The matched pairs (p,g) are mathematically proven to be *unique matching* (33) if their IoU(p,g)> 0.5. The *unique matching* splits the prediction and ground truth into three sets: the number of matched pairs (*TP*), the number of unmatched predictions (*FP*), and the number of unmatched ground truths (*FN*).

Metrics for Overlapped Nuclei Segmentation. We improved the Aggregated Jaccard Index (AJI) and accuracy metrics and proposed two new metrics to evaluate overlapped nuclei segmentation, namely, AJI of overlapped nuclei (AJIO), and accuracy for overlapped nuclei (ACCO). Because of the existence of many non-overlapped nuclei in images, traditional evaluation metrics cannot accurately validate the performance of overlapped nuclei segmentation. The proposed two metrics exclude all non-
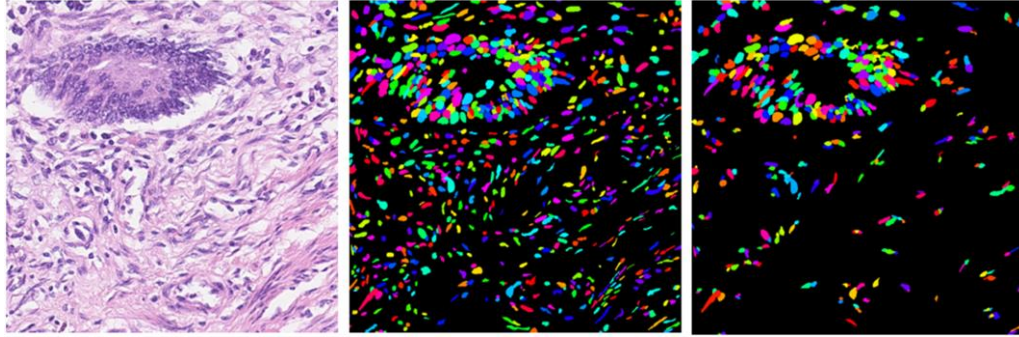
Figure 2.7 Ground truth of an example histopathological nuclei image (Graham, Vu, et al., 2019). From left to right: original image, ground truth of all nuclei, and ground truth of overlapped nuclei.

overlapped nuclei and focus on the evaluation of overlapped nuclei. Let $G = \{G_i\}_{i=1}^{N}$ be the overlapped nuclei in a ground truth image; and $S = \{S_k\}_{k=1}^{M}$ be the nuclei in the output image. AJIO is defined by

$$\text{AJIO} = \frac{\sum_{i=1}^{N} G_i \cap S_j}{\sum_{i=1}^{N} G_i \cup S_j}$$

where S_j is the matched nucleus in S that produces the largest Jaccard Index value with G_i.

Let M be the number of matched nuclei pairs between the segmentation and ground truth, and O denote the total number of overlapped nuclei in an image. For each overlapped nuclei, we iterate all the predicted segments and count two nuclei matched if their Jaccard Index value is larger than a threshold τ (0.5). The ACCO is given by

$$\text{ACCO} = \frac{M}{O}$$

The two metrics are general and can be applied to other overlapped object segmentation.

2.3.2 Implementation and Training

The proposed approach is trained by using an NVIDIA Titan Xp GPU. The encoder was pre-trained on ImageNet, and we trained the decoder for 100 epochs to obtain the initial parameters for the decoder branches. The network was further fine-tuned for 100 epochs on the nuclei training set. The size of the final output images is 80×80 pixels, and these output images are merged to form images with the same size (1000×1000) as the original images. In experiments, the initial learning rate is 10^{-4} and is reduced to 10^{-5} after 50 epochs. The batch size is 8 for training the decoder and 2 for fine-tuning the network. Moreover, processing an image of size 1000×1000 with our architecture takes about one second.

The input dimensionality of the network is 270×270×3. We prepare the training, validation, and test sets by extracting patches from images with 270×270 pixels size. During the training stage, data augmentation strategies, i.e., rotation, Gaussian blur, and median blur, are utilized for generating more

Table 2.1 Bend-Net: effectiveness of the proposed multi-task learning architecture using the CoNSeP dataset.

Methods	Metrics					
	AJI ↑	Dice ↑	RQ ↑	SQ ↑	PQ ↑	AJIO ↑
Instance-Net	0.371	0.841	0.603	0.771	0.471	0.296
HoVer-Net	0.545	0.840	0.674	0.773	0.522	0.520
Ours-OHV*	0.559	0.847	0.692	0.774	0.537	0.531
Ours-skip*	**0.565**	**0.850**	**0.697**	**0.779**	**0.544**	**0.537**

* Ours-OHV denotes the proposed approach with the OHV branch; Our-skip has additional skip connections between the HV and OHV branches.

images. The ground truths of overlapped nuclei are two or multiple individual nuclei have connected-component labeling. An example histopathology image, the ground truth of all nuclei, and the overlapped nuclei are demonstrated in Figure 2.7.

The proposed scheme comprises three stages: 1) preprocessing; 2) training of the proposed multi-task learning network; and 3) postprocessing. The preprocessing performs color normalization to reduce the impact of variations from the H&E staining and scanning processes (Vahadane et al., 2016). The postprocessing described in Graham, Vu, et al. (2019) is employed in this study, which applies Sobel operators to the distance maps to generate the initial contour map; then the difference between the initial nuclei contour map and nuclei instance map is used to generate markers; and finally, the watershed algorithm is applied to generate nuclei regions.

2.3.3 Effectiveness of the Network Architecture

The proposed multi-task learning architecture uses HoVer-Net as the backbone and integrates our newly proposed overlapped nuclei (OHV) branch and skip connections (Figure 2.6). To demonstrate the effectiveness of the proposed architecture, we compare the proposed network with the single-task network (Instance-Net), and two-task network (HoVer-Net). To perform a fair comparison, the proposed bending loss is not used. The approaches are evaluated on the CoNSeP dataset by using AJI, Dice, RQ, SQ, and PQ scores. As shown in Table 2.1, the Instance-Net does not apply any strategy to separate the overlapped nuclei and achieved very limited performance, e.g., AJI is only 0.371. The proposed network with the OHV branch ('Ours-OHV') achieved better average performance than the Instance-Net and HoVer-Net. With the new skip connections between the HV and OHV branches, the AJI, RQ, and PQ scores of the proposed approach ('Ours-skip') increased by 3.54%, 3.30%, and 4.04%, respectively. The AJIO scores demonstrate that the proposed approach outperforms HoVer-Net in separating overlapped nuclei.

Table 2.2 Bend-Net: effectiveness of the proposed bending loss using the CoNSeP dataset.

Methods	w/o bending loss	L_{be} v1	L_{be} v2	Metrics					
				AJI ↑	Dice ↑	RQ ↑	SQ ↑	PQ ↑	AJIO ↑
HoVer-Net	✓			0.545	0.840	0.674	0.773	0.522	0.520
		✓		0.552	0.844	0.683	0.774	0.530	0.523
			✓	0.559	0.846	0.690	0.776	0.537	0.528
Ours	✓			0.565	0.850	0.697	0.779	0.544	0.537
		✓		0.570	0.847	0.701	0.777	0.547	0.541
			✓	**0.578**	**0.851**	**0.709**	**0.781**	**0.555**	**0.552**

* L_{be} v1 and L_{be} v2 refer to our previous bending loss [35] and the newly proposed bending loss, respectively.

2.3.4 Effectiveness of the Bending Loss

The newly proposed bending loss improves our original bending loss calculation (Wang et al., 2020) by characterizing the difference between the concave and convex contour points. First, we compare the proposed multi-task learning network without any bending loss, with the bending loss (L_{be} v1) in (Wang et al., 2020), and with the newly proposed bending loss (L_{be} v2). Second, we demonstrate the effectiveness of the bending loss by adding it to HoVer-Net. The CoNSeP dataset and AJI, Dice, RQ, SQ, and PQ scores are used in experiments. As shown in Table 2.2, the proposed architecture with the v1 bending loss (Wang et al., 2020) achieves better performance than that of the network without any bending loss, and the proposed architecture with the newly proposed v2 bending loss outperforms the network with the v1 bending loss. The results demonstrate that the v2 bending loss can improve the overall performance (AJI: from 0.565 to 0.578) of nuclei segmentation. Meanwhile, adding the v1 or v2 bending losses to HoVer-Net improves its overall performance, which demonstrated the potential of applying the bending loss to improve the performance of other approaches. In addition, the AJIO scores demonstrate that the proposed approach with the v2 bending loss outperforms all other approaches in separating overlapped nuclei.

2.3.5 Parameter Tuning

Two hyper-parameters, α and μ, exist in the proposed loss function. α balances the bending loss and all other losses (Eq. 2.5); and μ (Eq. 2.3) gives different weights to the concave and convex contour points when calculating the bending loss. We conducted a grid search for the two parameters on the CoNSeP dataset by using the AJI score. Figure 2.8 shows the AJI results of nine parameter combinations (μ: 10, 20, 40; α: 0.5, 1.0, 2.0). As shown in Figure 2.8, the proposed approach achieved the best performance when μ is 20, and α is 1.0. Therefore, the bending loss of a concave curve pattern

Table 2.3 Bend-Net: overall test performance on the CoNSeP and MoNuSegv1 datasets.

Methods	CoNSeP					MoNuSegv1				
	AJI ↑	Dice ↑	RQ ↑	SQ ↑	PQ ↑	AJI ↑	Dice ↑	RQ ↑	SQ ↑	PQ ↑
FCN8	0.289	0.782	0.426	0.697	0.297	0.426	0.779	0.592	0.708	0.421
U-Net	0.482	0.719	0.490	0.668	0.328	0.520	0.722	0.635	0.675	0.431
SegNet	0.461	0.699	0.482	0.667	0.322	0.508	0.797	0.672	0.742	0.500
DCAN	0.408	0.748	0.492	0.697	0.342	0.515	0.778	0.659	0.718	0.473
DIST	0.489	0.788	0.500	0.723	0.363	0.560	0.793	0.618	0.724	0.449
Micro-Net	0.531	0.784	0.613	0.751	0.461	0.581	0.785	0.700	0.737	0.517
HoVer-Net	0.545	0.840	0.674	0.773	0.522	0.606	0.818	0.765	0.767	0.588
BEND	0.553	0.846	0.683	0.776	0.530	0.627	0.827	0.770	0.766	0.590
Bend-Net	**0.578**	**0.851**	**0.709**	**0.781**	**0.555**	**0.635**	**0.832**	**0.780**	**0.771**	**0.601**
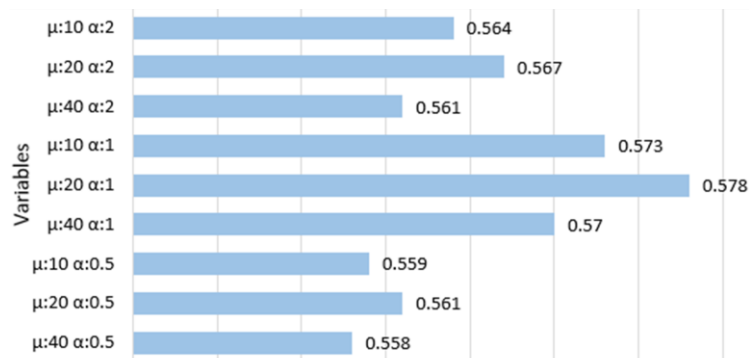
Figure 2.8 Fine-tuning parameters using AJI scores.

is twenty times the quantity of the same convex curve pattern. Refer to Figure 2.3 for the bending loss of different curve patterns.

2.3.6 Performance Comparison of State-of-the-art Approaches

We compared eight deep learning-based approaches, including three widely used biomedical segmentation architectures: FCN8 (Long et al., 2015), U-Net (Ronneberger et al., 2015), and SegNet (Badrinarayanan et al., 2017), and five state-of-the-art nuclei segmentation approaches: DCAN (Chen et al., 2016), DIST (Naylor et al., 2018), Micro-Net (Raza et al., 2019), HoVer-Net (Graham, Vu, et al., 2019), and BEND (Wang et al., 2020). Table 2.3 shows the overall performance of nine approaches on two public datasets (CoNSeP and MoNuSegv1) using five metrics AJI, Dice, RQ, SQ, and PQ. Note that all other approaches are tested using the described experiment settings, and therefore, the values in Table 2.3 may not be the same as those reported in the original publications. The Watershed algorithm is applied to FCN8, U-Net, and SegNet for postprocessing, whereas the rest of the approaches are implemented by following the same strategy as in the original papers. As shown in Table 2.3, the

Table 2.4 Bend-Net: overlapped nuclei segmentation performance on the CoNSeP and MoNuSegv1 Datasets.

	CoNSeP		MoNuSegv1	
Methods	AJIO ↑	ACCO ↑	AJIO ↑	ACCO ↑
FCN8	0.350	0.328	0.337	0.358
U-Net	0.486	0.395	0.472	0.464
SegNet	0.411	0.262	0.407	0.406
DCAN	0.417	0.293	0.427	0.423
DIST	0.542	0.476	0.543	0.536
Micro-Net	0.513	0.495	0.513	0.504
HoVer-Net	0.520	0.558	0.542	0.613
BEND	0.529	0.561	0.553	0.627
Bend-Net	**0.552**	**0.586**	**0.570**	**0.656**

proposed method outperforms other eight approaches in terms of all five metrics. Among three general biomedical segmentation architectures, U-Net achieved the highest AJI and RQ scores, but it has lower Dice and SQ scores than those of FCN8. DCAN and DIST built upon FCN8 and U-Net, respectively. DCAN outperforms the FCN8 in all five metrics, and DIST outperforms U-Net in all five metrics. However, their overall segmentation performances still are limited. Micro-Net, HoVer-Net achieve comparative segmentation results. The proposed Bend-Net achieves better results than all other approaches on the two datasets in all five metrics.

2.3.7 Overlapped Nuclei Segmentation

We proposed two new metrics, AJIO and ACCO, to evaluate overlapped nuclei segmentation. Table 2.4 shows the performance of overlapped nuclei segmentation on the CoNSeP and MoNuSegv1 datasets by using AJIO and ACCO scores. The DIST, Micro-Net, HoVer-Net, and the proposed method applied strategies to separate overlapped nuclei; therefore, their performances are significantly better than FCN8, U-Net, and SegNet. Our method achieved the best AJIO and ACCO scores on two datasets. Figure 2.9 shows segmentation examples of six image regions with overlapped nuclei from the CoNSeP and MoNuSegv1 datasets. In the ground truth images, overlapped nuclei are represented using different colors. In the result images, if an approach can separate two overlapped nuclei, the two nuclei should be in two different colors. As shown in Figure 2.9, FCN8, SegNet, and U-Net tend to miss small nuclei, and cannot separate overlapped nuclei. DCAN is slightly better at handling overlapped nuclei than FCN8, SegNet, and U-Net, but tends to miss nuclei that are not small. DIST separates overlapped nuclei better than the last four approaches, but it tends to over-segment nuclei and generate many small regions. Micro-Net performs well in segmenting overlapped nuclei, but it produces smaller nuclei

(a) Images (b) Ground truth (c) FCN8 (d) SegNet (e) U-Net (f) DCAN (g) DIST (h) Micro-Net (i) HoVer-Net (j)
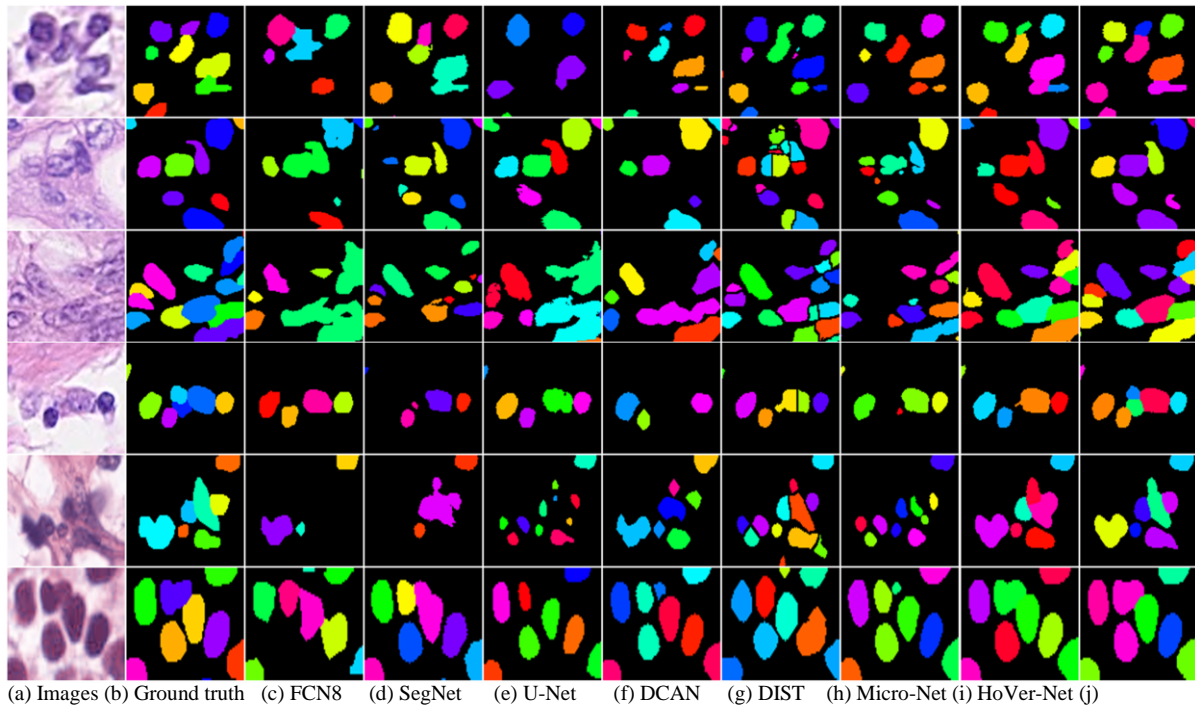
Figure 2.9 Samples of comparative segmentation results for state-of-the-art models.

regions than the ground truth and tends to over-segment nuclei. Hover-Net shows better segmentation results than the other six approaches. It can segment out small nuclei, and the sizes of result nuclei regions are close to those of the ground truth. It outperforms the other six approaches in segmenting overlapped nuclei; however, it has difficulty in separating closely touched nuclei. In Figure 2.9, the proposed method achieves the most accurate results on six images. The proposed approach not only can segment out small nuclei but also can separate closely touched nuclei accurately.

2.4 Bend-Net Conclusion

In this study, we propose a novel deep multi-task learning network to address the challenge of segmenting overlapped nuclei in histopathology images. Firstly, we propose the bending loss regularizer, which defines different losses for the concave and convex points of nuclei contours. Experimental results demonstrated that the bending loss effectively improves the overall performance of nuclei segmentation, and it can also be integrated into other deep learning-based segmentation approaches. Secondly, the proposed multi-task learning network integrates the OHV branch to learn knowledge from the overlapped nuclei, which enhances the segmentation of touching nuclei. Thirdly, we proposed two quantitative metrics, AJIO and ACCO, to evaluate overlapped nuclei segmentation. The extensive experimental results on two public datasets demonstrate that the proposed Bend-Net achieves state-of-the-art performance for nuclei segmentation. In the future, we will extend the

proposed approach to more challenging tasks, such as gland segmentation and semantic image segmentation.

Chapter 3: Topology-Aware Network for Histopathology Gland Segmentation

Wang, H., Xian, M., & Vakanski, A. (2022). TA-Net: Topology-aware network for gland segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.

3.1 Introduction

In histopathology image analysis, the assessment of gland morphology is crucial for determining the stages of several cancers, e.g., colon cancer (Fleming et al., 2012), breast cancer (Bloom & Richardson, 1957), and prostate cancer (Montironi et al., 2005). Accurate gland segmentation can aid in identifying various histopathological features, such as glandular shape, size, distribution, and density, which can be useful for cancer diagnosis, grading, and prognosis. Traditionally, pathologists examine gland morphology to assess the malignancy degree using microscopes, which is a time-consuming, expensive, and error-prone process. Recently, with the availability of whole slide images (WSI), digital pathology has gained popularity by developing computational tools to aid routine tasks. Automatic and accurate gland segmentation is often required before calculating gland morphology. However, this task is challenging due to the large morphological differences among glands and the large number of clustered glands.

Early approaches for gland segmentation focused on applying prior knowledge of glandular structures, such as morphology-based methods (Nguyen et al., 2010; Paul & Mukherjee, 2016), and graph-based methods (Tosun & Gunduz-Demir, 2010). These methods achieved promising performance on low-grade adenocarcinoma. However, these methods failed to accurately segment malignant glands, which tend to grow, invade adjacent tissues, and cluster densely, causing severe deformation of their shapes in histopathology images (as shown in Figure 3.1). Recently, deep learning-based methods provide state-of-the-art performance in many computer vision tasks (Badrinarayanan et al., 2017; Chen et al., 2017) and biomedical image analysis tasks (Ronneberger et al., 2015). For instance, Chen et al. (2016) proposed a DCAN that has FCN-based multi-task learning network to generate gland instances and contours simultaneously. The complementary contour information helped separate clustered glands. Xu et al. (2017) developed a deep three-channel network to jointly separate the clustered glands using instance, contour, and location information. Raza et al. (2019) proposed the Micro-Net that inputs the multiple resolutions of image patches at different down-sampling stages for
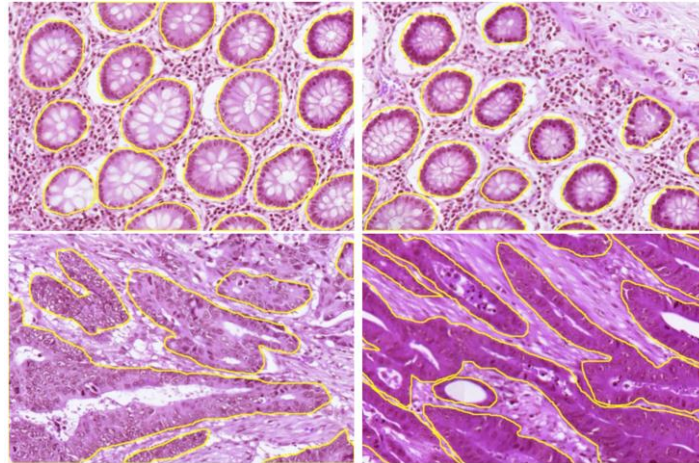
Figure 3.1 Hematoxylin and Eosin (H&E) stained histopathology images with labeled (yellow contours) glands (Sirinukunwattana et al., 2015). The first row shows healthy glands, the second row shows malignant glands. Noted that the malignant glands are close their their neighboring glands and appear in deformed shape.

(a) Image patch   (b) Ground truth   (c) DCAN

(d) SegNet   (e) MicroNet   (f) TA-Net
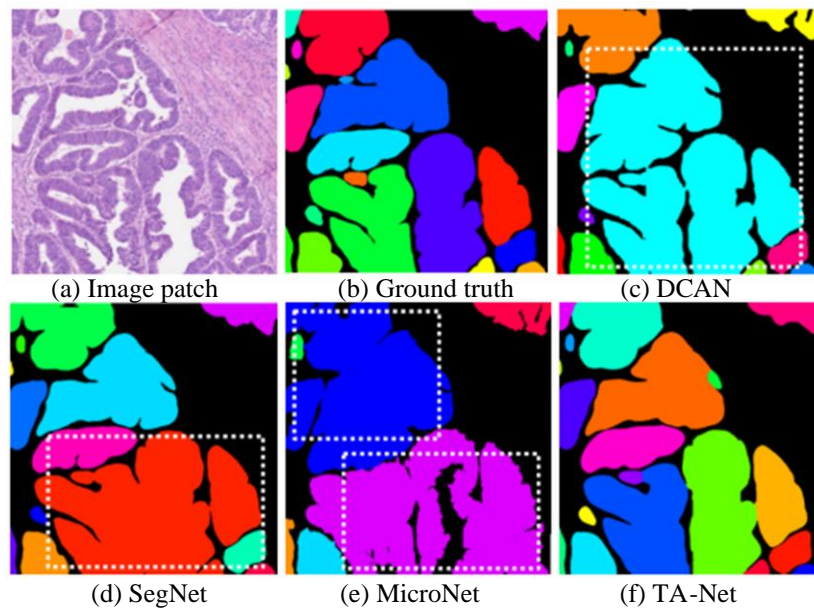
Figure 3.2 A segmentation example of densely clustered glands. Colors are used to differentiate different glands. The white dash rectangles highlight the poorly-separated glands.

better localization and context information and back-propagates the results by using multi-resolution outputs. Qu et al. (2019) proposed a full-resolutional network that outputs three-class probability maps (instance, contour and background). The strategy shared by all the above methods is to use gland contour information to separate clustered glands. However, these approaches achieved limited success. A segmentation example of densely clustered glands is shown in Figure 3.2, SegNet (Badrinarayanan et al., 2017), DCAN (Chen et al., 2016), and Micro-Net (Raza et al., 2019) failed to separate close glands.
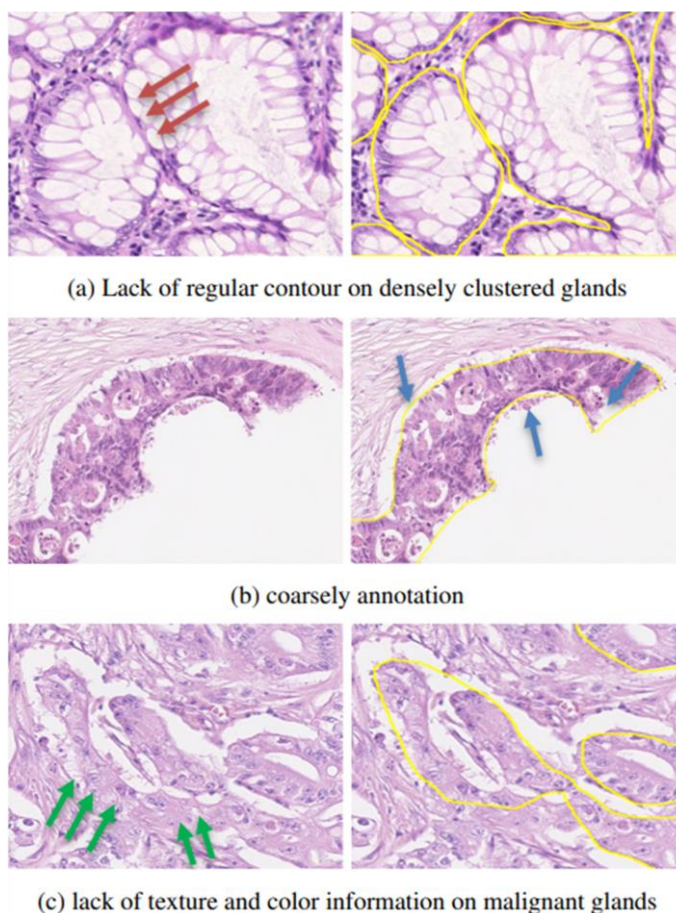
Figure 3.3 Issues of segmenting densely clustered glands by using a gland contour annotation. Left column shows the histopathology image patches; right column shows the image patches with the labeled gland contours (yellow).

Three major challenges exist in gland segmentation using contour information. 1) The contour strategy fails when glands are densely clustered because overlapped glands share contour sections. Usually, in a glandular structure, epithelial nuclei form a gland border. In practice, a digital WSI scanner flattens gland tissues into a near two-dimensional histopathology slice, and two or multiple glands cluster together will not have a regular and clear epithelial nuclei border. Note the red arrow in Figure 3.3 (a), the clustered regions do not have regular epithelial nuclei. 2) The coarse annotations of the contours introduce noise and reduce the effectiveness of the contour strategy. A gland tissue in a $20 \times$ magnification could be $1k \times 1k$ pixels wide and height, and it is difficult to annotate all contour pixels correctly. Existing datasets in the literature still have many annotation issues (Figure 3.3 (b)). 3) It is difficult to identify the contours of malignant glands accurately. Because malignant glands continue to grow and become deformed. their components appear distorted. The green arrows in Figure 3.3 (c) indicate the distorted boundary of a malignant gland. In this study, we introduce a novel approach that utilizes gland topology to separate densely clustered glands. The gland topology is characterized by the

topological skeleton of glands, which differentiates clustered glands more accurately than gland contours. Furthermore, the gland topology is more robust than contours in the presence of noisy annotations. This work has the following two major contributions. 1) We propose a topology-aware network that can simultaneously perform instance segmentation and gland topology estimation. The topology branch of the network predicts the Medial Axis (Blum, 1967) distance map, which provides an accurate description of gland topology. 2) We propose a new topology loss that utilizes the Medial Axis distance map and gland markers. This loss penalizes the topological difference between segmented glands and the true glands, forcing the network to generate segmentation results that adhere to the gland topology.

3.2 Related Works

Histopathology Gland Segmentation. Histopathology gland segmentation aims to segment the gland tissue from the Hematoxylin & Eosin (H&E) stained histopathology image. Recently, deep learning (DL)-based method successfully demonstrates the robustness and efficiency in the literature. Graham, Chen, et al. (2019) proposed the MILD-Net, which utilized both instance and contour segmentation, as well as multi-level aggregation, atrous spatial pyramid pooling block, and dilated convolutional design. Ding et al. (2020) proposed a multi-scale Fully convolutional network to extract different receptive field features at different convolutional layers. These studies, as well as those described in the previous section (Chen et al., 2016; Qu et al., 2019; Raza et al., 2019; Xu et al., 2017), build up a deep architecture for segmenting the gland instance and contours. Yan et al. (2020) proposed a shape-aware adversarial learning network that integrates a deep adversarial network and a shape-preserve loss. Qu et al. (2019) proposed a spatial loss for recognizing the glands. The proposed loss placed a spatial constraint on the boundary pixels and forced the network to learn gland shapes.

Topology Aware Network. Different Topology aware networks have been proposed in various natural image segmentation (Clough et al., 2020; Estrada et al., 2014; Mosinska et al., 2018) and biomedical image segmentation tasks (Hu et al., 2019; Shit et al., 2021). Hu et al. (2019) introduced a loss function that makes the segments have the same Betti number as the ground truths for the topological correctness; and the method utilized the topology information to make the corrections on some biological structures, e.g., broken connection. Clough et al. (2020) proposed a method that integrated the differentiable properties of persistent homology into the network training process; the network extracts useful gradients even without ground truth labels. Shit et al. (2021) introduced a centerline Dice that measured the topological similarity of the segmentation masks and their skeleton. Mosinska et al. (2018) constructed a loss that models higher-order topological information. These methods employed regional topological constraints, e.g., connectivity and loop-freeness. However, these
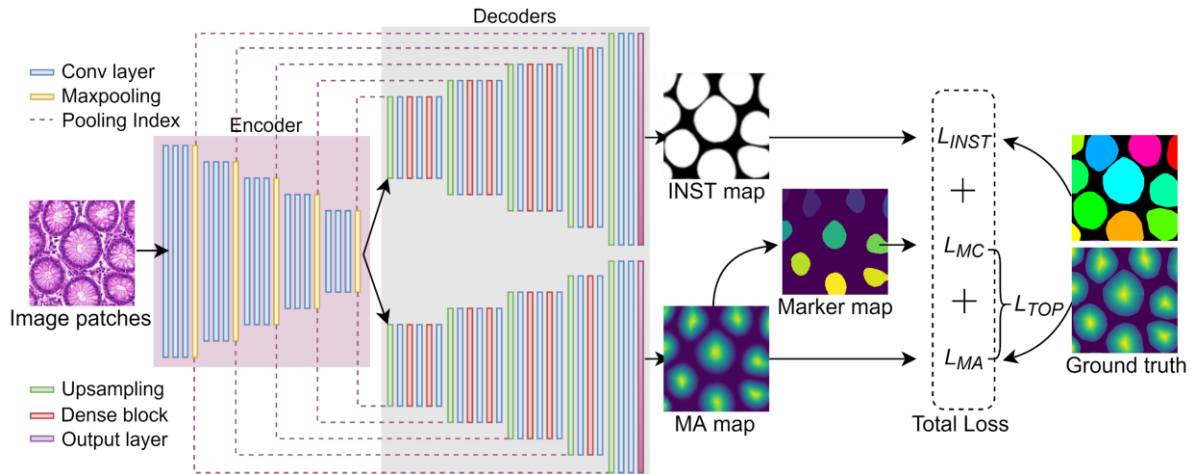
Figure 3.4 The architecture of the proposed network. The architecture takes image patches as input and outputs the gland instance (INST) map and medial axis (MA) distance map. The marker map is generated using the MA map.

methods are hard to generalize to other objects without linear structures, e.g., blood vessels, retinas, cracks, and roots. In this work, we introduce a new topology-aware network to preserve the topological skeletons of glands, which can be easily reproduced in objects with irregular shapes and overlapping issues.

Medial Axis (MA) Transform. The MA/topological skeleton of an object is the set of all inner points that have more than one closest point on the object's boundary. The MA transformation was first introduced by Blum (1967) for shape recognition, and it is well-known as the locus of local maxima with distance transformation. The MA transform is a powerful tool for shape abstraction, and provides a shape representation that preserves the topological property of skeleton structures; these properties are invariant to crop, rotation, and articulation and robust to overlapped and clustered objects. Recently, many studies applied medial axis transforms in different computer vision tasks, e.g. segmentation (Noble & Dawant, 2011), shape matching (Siddiqi et al., 1998), recognition (Sebastian et al., 2001), image reconstruction (Tsogkas & Dickinson, 2017), and body tracking (Shotton et al., 2012). In a densely clustered gland region, different glands may contain different skeletons. The skeleton structure may assist to distinguish the densely clustered glands. However, to the best of our knowledge, no studies have employed MA transform in histopathology image analysis.

3.3 Topology-Aware Network for Gland Segmentation

Figure 3.4 illustrates the architecture of the proposed TA-Net. It is a deep multi-task neural network designed to enhance gland segmentation by simultaneously learning shared representation from two tasks: gland instance segmentation (INST) and gland topology estimation (TOP). The INST task

extracts glands from the background. The second task learns gland topology to separate clustered glands and segment the glands accurately.

3.3.1 TA-Net Architecture

The proposed TA-Net has one encoder and two decoder branches. The first decoder predicts the foreground map for glands (INST), and the second decoder learns the topology information of glands (TOP). Two decoders share the same feature maps from the encoder. SegNet (Badrinarayanan et al., 2017) is utilized as the backbone architecture. Meanwhile, dense-connected blocks (Gao Huang, 2017) are applied to the decoders to ensure a large receptive field for detecting the instances over wider areas in images. In the proposed network, the two decoders have the same architectures except for the final output layers. The INST decoder ends with a 2 by 2 convolutional layer and follows a softmax activation layer, and the TOP decoder ends with a 1 by 1 convolutional layer for outputting the medial axis distance map. In the encoder, three convolutional layers and the following max-pooling layer form a downsampling convolutional block. In total, there are five downsampling blocks. In the decoder, there are five upsampling blocks that contain different numbers of stacked densely connected layers and convolutional layers. Different from the standard SegNet encoder architecture, there have three convolutional layers in the first two downsampling blocks that aim at extracting more fundamental features. In TA-Net, all convolutional kernels are 3×3, and the numbers of kernels for the convolutional layers in each block are the same. The numbers of convolutional kernels from blocks 1 to 5 in the encoder are 64, 128, 256, 512, and 512, respectively. In the two decoders, the numbers of kernels from blocks 1 to 5 are 512, 512, 256, 128, and 64, respectively; and the numbers of stacked densely-connected layers are 8, 8, 8, 8, 4, 4, and 4, respectively. The loss function of TA-Net. As shown in Figure 3.4, the proposed TA-Net's loss function has two terms: the instance loss ($\mathcal{L}_{INST}$) and the topology loss ($\mathcal{L}_{TOP}$). The total is defined by

$$\mathcal{L}_{TA-Net} = \mathcal{L}_{INST} + \alpha \cdot \mathcal{L}_{TOP} \tag{3.1}$$

where α denotes the contribution of the topology loss. The loss $\mathcal{L}_{INST}$ is the cross-entropy (CE) loss on gland instances map for segmenting the foreground gland instances from the background. The loss $\mathcal{L}_{TOP}$ is discussed in Section 3.4.2, α controls the contribution of the $\mathcal{L}_{TOP}$ loss.

3.3.2 Topology Loss

The proposed topology loss is given by

$$\mathcal{L}_{TOP} = \mathcal{L}_{MA} + \mathcal{L}_{MC} \tag{3.2}$$

(a)	(b)	(c)	(d)
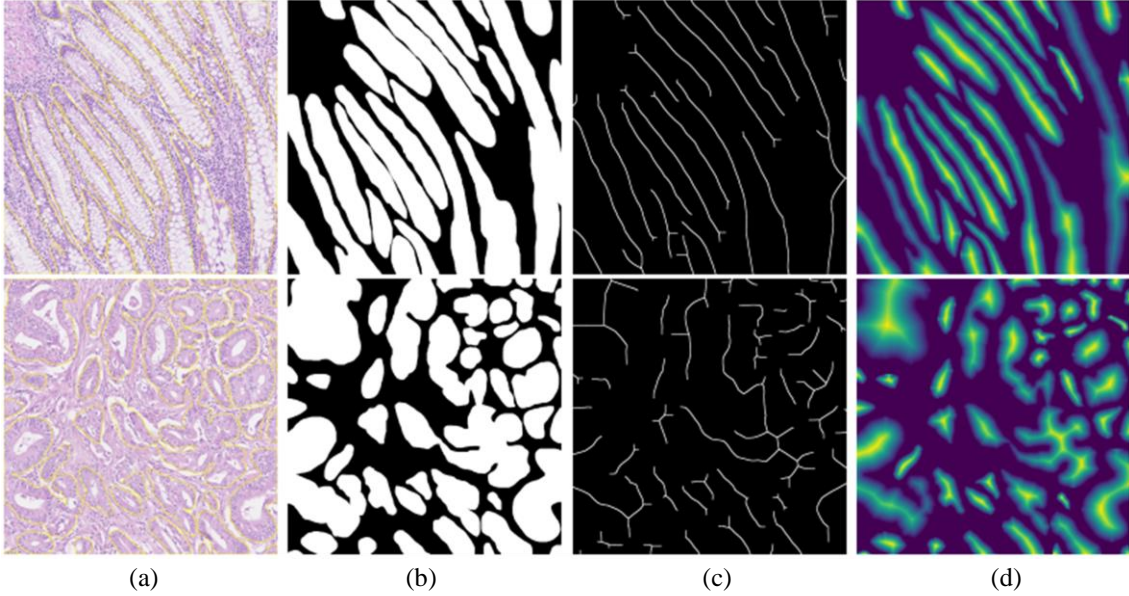
Figure 3.5 Examples of Medial axis (MA) transformation. a) A histopathology image patch with labeled (yellow contours) clustered glands; b) the binary annotation of gland regions; c) the topological skeletons of glands, noted that skeletons are morphological dilated to be visible; and d) the MA distance map.

where $\mathcal{L}_{MA}$ is computed using the medial axis distance map to preserve the geometry of glands, and $\mathcal{L}_{MC}$ uses markers in glands to avoid over-segmentation and under-segmentation. Medial Axis (MA) Distance Map. The MA/topological skeleton of an object is the set of all inner points that have more than one closest point on the object's boundary. It is well-known as the locus of local maxima with distance transformation. In this work, we employ the MA-based distance map to model the topological property of glands. Let $G = \{G_i\}_{i=1}^{n}$ be a set of glands in an image patch, and n be the number of glands. For every gland from G_1 to G_n, the MA transformation iterates the one-pixel morphological erosion process starting through the gland contour. The topological skeleton of a gland (Figure 3.5(c)) is a set of points having more than one closest point on a gland's contour, and the skeletons are one-pixel width and follow the same connectivity as the original gland shape. The number of iterations from a gland's contour to the topological skeleton is normalized to form the MA distance map. The MA distance map value at point p_j is defined by

$$MA(p_j) = \begin{cases} \frac{d(p_j)}{\max_{\forall p_k \in G_i}\{d(p_k)\} - \min_{\forall p_k \in G_i}\{d(p_k)\}}, & if \exists G_i \in G, p_j \in G_i \\ 0, & \text{otherwise} \end{cases}$$

(3.3)

where $d(p_j)$ is the number of erosion iterations from point p_j to the corresponding skeleton. Examples of the MA distance map is shown in Figure 3.5 (d). As shown in Figure 3.5 (a), it is challenging to separate the clustered glands; however, in Figure 3.5 (c) and (d), the skeleton and distance map

emphasizes the geometrical and topological properties of each gland and clearly separates all clustered glands. Eq. 3.3 and the MA transformation are applied to generate the ground truth for the MA distance map. To ensure that the proposed gland segmentation network preserves gland topology, we design the second decoder branch to predict the MA distance map. The loss function $\mathcal{L}_{MA}$ is defined by

$$L_{MA} = \frac{1}{m} \sum_{j=1}^{m} (MA(p_j) - \widehat{MA}(p_j))^2$$

(3.4)

where $\widehat{MA}$ denotes the predicted distance map, and m is the number of image pixels.

Marker loss. The Watershed algorithm is commonly applied as a postprocessing step to produce fine segmentation, especially separating the clustered objects. The watershed markers represent the number and locations of objects and are critical for accurate segmentation. More markers lead to over-segmentation, and fewer markers produce under-segmented results. We introduced the marker loss to separate clustered glands and prevent over-segmentation. The marker loss is defined as the Dice loss between the predicted marker map ($\widehat{MC}$) and the true marker map (MC)

$$L_{MC} = Dice(MC, \widehat{MC})$$

(3.5)

the predicted marker map is generated by thresholding the outputs from the medial axis distance map.

3.4 Experimental Results

3.4.1 Datasets and Evaluation Metrics

Datasets. The Colorectal adenocarcinoma gland (CRAG) dataset (Graham, Chen, et al., 2019) and the Gland Segmentation challenge (GlaS) dataset (Sirinukunwattana et al., 2015) are used in this work. CRAG has 213 H&E-stained histopathology images from 38 WSI images. The scanned image size is 1512×1516 pixels with the corresponding instance-level ground truth. The training set has 173 images and the test set has 40 images with different cancer grades. The GlaS dataset has 165 H&E-stained histopathology images extracted from 16 WSI images. The image size mostly is 775×522 pixels. The training set has 85 images (37 benign and 48 malignant). The test set is split into two sets: Test A (60 images) and Test B (20 images), because two test sets are released in different stages in GlaS challenge. Both datasets are scanned with a 20× objective magnification. The CRAG dataset has more densely clustered glands.

    Evaluation Metrics. We use the F1-score, object-level Dice coefficient (Obj-D), and object-level Hausdorff distance (Obj-H). In the F1 score, a segmented gland is counted as a true positive if it has >50% overlap with the ground truth, and counted as a false positive (FP) if otherwise; all missed glands

in the ground truth are counted as false negatives. Refer to Sirinukunwattana et al. (2017) for detailed descriptions of the Obj-D and Obj-H in the Colon Histology Images Challenge Contest (GlaS) at MICCAI 2015.

3.4.2 Implementation Details

The TA-Net is trained and tested using a deep learning server with an NVIDIA Quadro RTX 8000 GPU, 512 GB memory, and two 2.4 GHz Intel Xeon 4210R CPUs.

The patch size of the GlaS dataset is 512×512 pixels, and the patch size of the CRAG dataset is 768×768 pixels. Different patch sizes are applied because 1) the two datasets have images with different sizes, and 2) larger patches improve the performance of segmenting white lumen regions inside the gland tissues. The larger patches could generate the whole white lumen region in the gland in one patch. The GlaS dataset generates 340 training patches and 320 test patches. The CRAG dataset generates 692 training patches and 160 test patches. The augmentation approaches, e.g., random flip, random rotation, Gaussian blur, and median blur, are utilized in the training stage. The segmentation results of image patches are merged to form images of the same size as the original images.

The training epoch is set as 200, and the initial learning rate for the Adam optimizer is set as 10^{-4} and is reduced to 10^{-5} after 100 epochs. The batch size is 4 for training the model.

The postprocessing applies the Watershed algorithm to produce the final output. We apply a threshold value to the outputs from the instance branch (INST) to generate the glands binary map and utilize it as the Watershed filling region. The output from the MA distance branch is the local elevation of those glands. Further, thresholding the MA distance map to generate the Watershed markers. Morphology operations, e.g., fill the holes, remove the small objects are utilized to generate fine glands regions and makers. In the end, the generated gland region, gland elevation, and markers are input to the Watershed algorithm for fine gland segmentation results.

Table 3.1 TA-Net: overall segmentation performance on GlaS and CRAG datasets.

Datasets	Methods	F1(%)↑	Obj-D(%)↑	Obj-H↓
GlaS	SegNet	83.1	84.9	76.6
	DCAN	81.4	83.9	102.9
	DeepLab	83.7	84.5	80.5
	MILD-Net	87.9	87.5	73.7
	Micro-Net	86.5	87.6	70.4
	FullNet	88.9	88.5	63.0
	DSE	89.4	89.9	55.9
	MSFCN	89.3	89.9	53.1
	Yan *et al.*	**90.7**	89.3	58.7
	TA-Net	90.5	**90.2**	**50.8**
CRAG	SegNet	77.4	85.3	134.7
	DCAN	73.6	79.4	218.8
	DeepLab	64.8	74.5	281.4
	MILD-Net	82.5	87.5	160.1
	DSE	83.5	88.9	120.1
	MSFCN	82.5	89.2	130.4
	TA-Net	**84.2**	**89.3**	**105.2**

3.5 Results and Discussion

In this section, we discuss the overall performance of our method, followed by the results using the contour map, and single-task/multi-task networks. Finally, we discuss the performance of our network on different distance metrics.

3.5.1 Overall performance

We compare the proposed method with nine recently published approaches using the GlaS dataset and six approaches using the CRAG datasets. We implemented SegNet (Badrinarayanan et al., 2017), DCAN (Chen et al., 2016) DeepLab (Chen et al., 2017) by following the same strategies in the original papers and MILD-Net (Graham, Chen, et al., 2019), Micro-Net (Raza et al., 2019), FullNet (Qu et al., 2019), DSE (Xie et al., 2021), MSFCN (Ding et al., 2020), Yan et al. (2020) are cited from their original papers. The test performance of the GlaS dataset is reported as the average performance on its test A and test B sets. We employed F1-score, object-level Dice coefficient (Obj-D), and object-level Hausdroff (Obj-H) distance to measure the overall performance. Table 3.1 shows the test performance of different approaches on two public datasets. The proposed TA-Net outperforms all other methods on CRAG datasets in terms of the F1 score, Obj-D; and achieved the best results in Obj-D, Obj-H, and the second-best results in F1-score on GlaS datasets. The GlaS dataset has a small number of densely clustered gland regions; therefore, compared to the second-highest result (MSFCN), TA-Net is only slightly better, e.g. 4.3% improvement of the Obj-H. The CRAG dataset has more densely clustered

| (a) Original Images | (b) DCAN | (c) Ours-CNT | (d) TA-Net | (e) Ground truth |
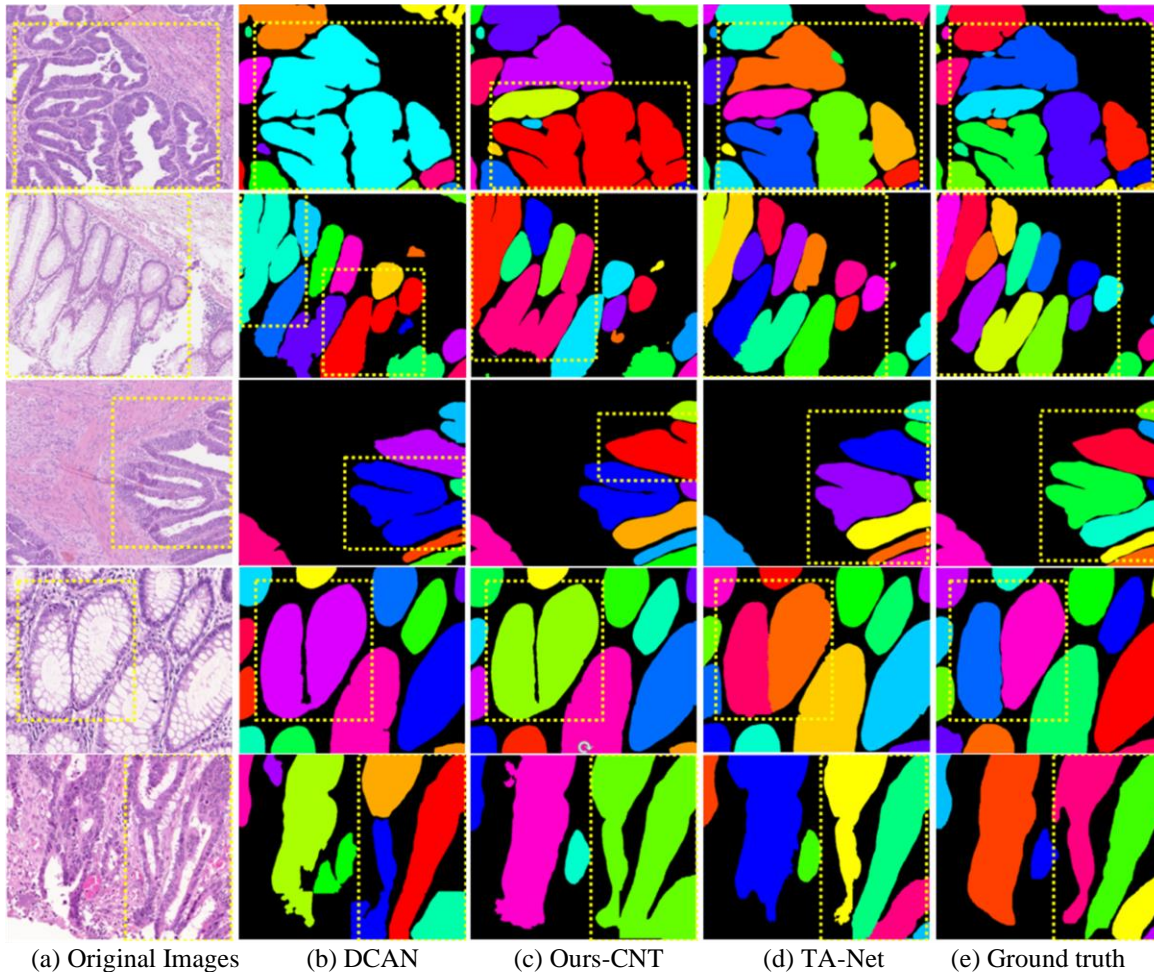
Figure 3.6 Segmentation results of five image patches (top three from CRAG, bottom two from GlaS). Different colors represent different glands. Yellow dash region highlighted the clustered gland regions.

glands, and our method improves the Obj-H significantly (19.3%). Figure 3.6 (d) demonstrates that TA-Net separates densely clustered glands accurately.

3.5.2 Effectiveness of Medial Axis Distance Map

We compared the proposed TA-Net with a multi-task network (Ours-CNT) which has the same architecture as TA-Net but outputs the gland instance map and the gland contour map. The only difference between the two networks is that TA-Net uses the MA distance map, while Ours-CNT uses the gland contour map as the ground truth of the second decoder. As shown in Table 3.2, TA-Net compared to the Ours-CNT, ObjH has been improved by 5.7% and 35% on the GlaS and CRAG datasets, respectively. The results demonstrate that the network using the MA distance map generates more accurate gland contours than the contour map-based network, especially on the CRAG dataset.

Table 3.2 TA-Net: ablation study on multi-task learning and decoders

Datasets	Methods	F1(%)↑	Obj-D(%)↑	Obj-H↓
GlaS	Ours-INST	86.4	88.3	65.4
	Ours-MA	80.2	84.2	105.8
	Ours-CNT	89.1	88.2	54.4
	TA-Net	**90.5**	**90.2**	**50.8**
CRAG	Ours-INST	78.9	86.1	125.6
	Ours-MA	74.8	80.3	200.8
	Ours-CNT	81.3	85.8	164.5
	TA-Net	**84.2**	**89.3**	**105.2**

Figure 3.6 demonstrates that the contour map-based strategy fails to separate many clustered glands (in dashed rectangles).

3.5.3 Effectiveness of Multi-task network

We compared the proposed multi-task learning network with the singl e-task learning network, which outputs the binary instance branch only (Ours-INST), and MA distance bra nch only (Ours-MA) . In the experiment of the Ours-INST, the Watershed algorithm is applied in the postprocessing for separating the clustered glands. In the experiment of MA distance branch, we output the MA distance branch only. In the post-processing, we utilized thresholding to produce a binary gland region map and gland markers. Then, the Watershed is used to produce the final segmentation. From Table 3.2, we noted that the designed multi-task learning network outperforms one task networks (Ours-INST, Ours-MA). Integrating both gland instance and MA distance map will produce a reliable performance in gland segmentation.

3.5.4 Effectiveness of the Marker Loss

The proposed TA-Net is compared with the same network without the marker loss (Ours-WoM). Table 3.3 shows the results on two public datasets. From the quantitative results, we noted that the marker loss only improves the overall performance slightly on both two datasets. But we observed that the marker loss alleviates the over-segmentation and under-segmentation problems in clustered glands and deformed glands in many qualitative cases (Figure 3.7).

Table 3.3 TA-Net: ablation study on marker loss

Datasets	Methods	F1(%)↑	Obj-D(%)↑	Obj-H↓
GlaS	Ours-WoM	90.2	89.8	54.4
	TA-Net	**90.5**	**90.2**	**50.8**
CRAG	Ours-WoM	83.7	89.1	108.6
	TA-Net	**84.2**	**89.3**	**105.2**

| (a) Images patches | (b) Ground truth | (c) Ours-WoM | (d) TA-Net |
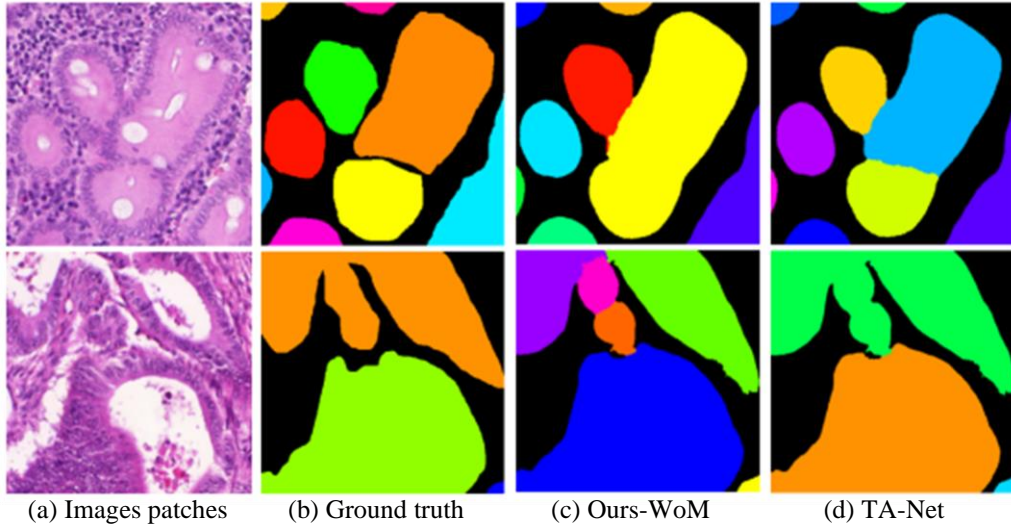
Figure 3.7 Examples of the effectiveness of the marker loss.

Table 3.4 TA-Net: ablation study on different distance metrics on CRAG dataset

Datasets	Metrics	F1(%)↑	Obj-D(%)↑	Obj-H↓
GlaS	Euclidean	82.4	78.7	65.4
	Chessboard	86.4	83.4	71.2
	MA	**90.2**	**89.8**	**54.4**
CRAG	Euclidean	76.5	81.2	257.5
	Chessboard	80.8	85.9	178.9
	MA	**83.7**	**89.1**	**108.6**

3.5.5 Comparison on Various Distance Maps

Deep Watershed-based regression algorithms provide a successful demonstration to separate the occluded objects and overlapped objects (Bai & Urtasun, 2017). The most related method to ours is by Naylor et al. (2018), which proposed a chessboard distance-based deep regression network for nuclei segmentation. First, compared to their U-Net shape architecture, we employed the multi-task learning-based Densely connected SegNet. We achieved promising performance in segmenting the gland foreground from the background. Second, the Medial Axis distance map preserves the topological property of the objects, which maintains the gland structure information during the training stage. Third, the marker loss will control the over-segmentation and under-segmentation issues for accurate marker detection. To demonstrate the effectiveness of MA distance transform with other distance-based metrics, we set up an experiment using our network test with different distance metrics, including the Euclidean distance, Chessboard distance, and Medial axis (MA) distance. To conduct a fair comparison, marker loss will not apply in this study, and we replace the MA distance metrics to other distance metrics. Similar to our post-processing approach, the predicted gland foreground segmentation and the

predicted distance map is utilized to produce the final fine segmentation. From the results in Table 3.4, we noted that our methods outperform the other two distance metrics. It gives the fact that medial axis distance metrics achieved the best performance comparing to use Euclidean and chessboard distance metrics.

3.6 TA-Net Conclusion

In this study, we propose a topology-aware network (TA-Net) to address the challenge of partitioning densely clustered glands in histopathology images. Firstly, the proposed multi-task learning architecture integrates both instance segmentation and gland topology learning and learns their shared representation. Experimental results show that TA-Net outperforms the state-of-art multi-task architectures, e.g., DCAN, and single-task architectures. Secondly, we propose a topology loss using a Medial Axis distance map and gland makers. The loss penalizes the topology changes between the segmented glands and the actual glands. The extensive experimental results on two public datasets demonstrate that the proposed TA-Net achieves state-of-the-art performance for densely clustered gland segmentation. In the future, we will extend the proposed approach to other challenging tasks, such as nuclei segmentation and semantic image segmentation.

Chapter 4: Style-Guided Instance-Adaptive Normalization for Multi-Organ Histopathology Image Synthesis

Wang, H., Xian, M., Vakanski, A., & Shareef, B. (2023). SIAN: Style-Guided Instance-Adaptive Normalization for Multi-Organ Histopathology Image Synthesis. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI).

4.1 Introduction
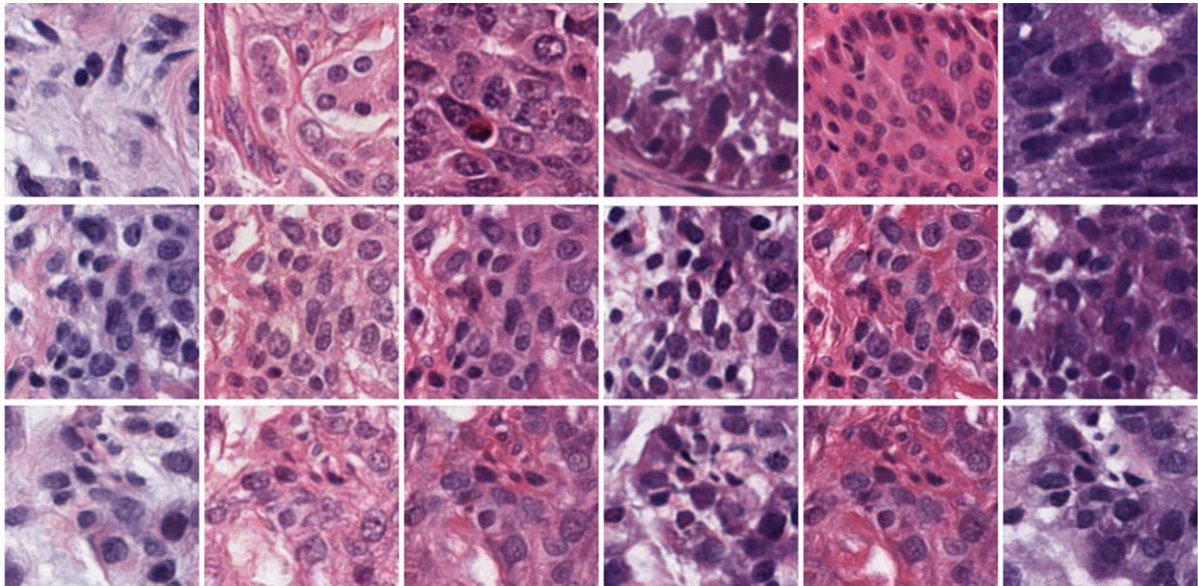
Figure 4.1 Examples of image synthesis for multiple organs. The top row shows real histopathology images from six organs. The second and third rows are synthesized images generated using the proposed approach (SIAN).

Histopathology image analysis has achieved great success in automatic tissue segmentation (Graham, Vu, et al., 2019; Wang et al., 2020) and cancer grading (Shaban et al., 2020). Existing deep learning-based methods require large fully-annotated datasets during the training stage, but current annotated datasets are relatively small. For example, only tens of image patches were used in Graham, Vu, et al. (2019); Kumar et al. (2017); Naylor et al. (2018); Vu et al. (2019). With large annotated datasets, we could train more accurate and reliable models. However, it is expensive to annotate large datasets for histopathology images, because each image may contain more than tens of thousands of nuclei.

To overcome the challenge, image synthesis is adopted. Recent works have demonstrated that high-quality synthetic images could improve the overall performance in histopathology image analysis (Butte et al., 2023; Butte et al., 2022; Deshpande et al., 2022; Wei et al., 2019; Xue et al., 2021).

However, these methods generated images only for single cancer or cancers with shared similarity, e.g., colorectal cancer (Deshpande et al., 2022; Wei et al., 2019), lymph node (Xue et al., 2021); thus, their models cannot generate different image styles for different cancer types. In practice, H&E-stained images for cancers from different organs could have large color and texture variances both in foreground nuclei and background stroma (first row of Figure 4.1). Therefore, it is important to have the network to generate histopathology images well in various stain distribution across multiple organs.

Recently, neural style transfer (NSF) methods have been widely exploited in many natural image synthesis tasks for manipulating image styles (Huang & Belongie, 2017; Karras et al., 2019; Park et al., 2019), they aim to learn the style from a reference image and apply it to the target image. SPADE (Park et al., 2019) extended the AdaIN norm (Huang & Belongie, 2017) into the spatially adaptive manner for obtaining semantic alignments and used the encoded style vector at the beginning of a network, which enabled simultaneously style manipulation and semantic image synthesis. However, most existing histopathology image synthesis methods applied semantic layouts as the network input to learn object-level image appearance (Deshpande et al., 2022; Park et al., 2019; Wei et al., 2019). In histopathology image analysis, a large amount of clustered and overlapped objects may have the same semantic class label, which makes it difficult to generate accurate boundaries among clustered objects.

To alleviate the above issues, we proposed a style-guided instance-adaptive normalization (SIAN) to combine the image style vector with the instance layout for modulating the GAN generator. The learned transformation can effectively propagate the network to learn style factors for synthesizing histopathology images across various color distributions and instance features for generating accurate densely-clustered nuclei. SIAN allows the user to choose a style image from a specific organ and synthesizes histopathology images with a similar style.

4.2 Related Work

Conditional Image Synthesis. Conditional generative adversarial networks (cGANs) engage additional information to guide image generation, and they have achieved promising results in many image synthesis tasks, e.g., image-to-image synthesis (Isola et al., 2017), multi-modal synthesis (Park et al., 2019), high-resolution image synthesis (T.-C. Wang et al., 2018), image-guided image synthesis (P. Wang et al., 2021). Pix2Pix GAN (Isola et al., 2017) learned mapping from an input image to an output image using the U-Net architecture. SPADE (Park et al., 2019) adopted a semantic layout for preserving the semantic class features and improved unconditional normalization methods.

Histopathology Image Synthesis. Histopathology image synthesis provides a promising solution for data-efficient methods without requiring extensive human resources. It has been widely studied in many applications (Deshpande et al., 2022; Hou et al., 2019; Wei et al., 2019; Xue et al., 2021). Deshpande
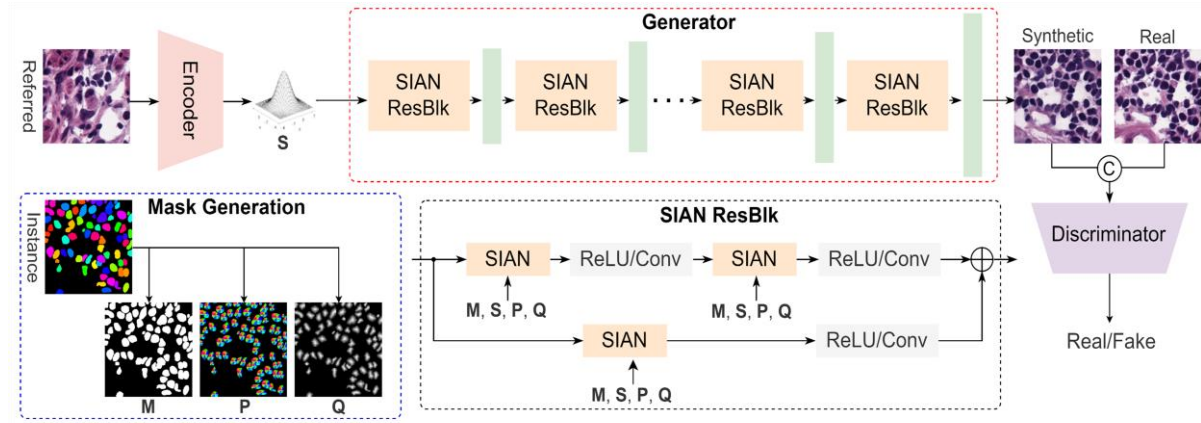
Figure 4.2 Architecture of the proposed method. The encoder learns style vectors from a referred image; and the SIAN blocks integrate image style (S), semantic map (M), directional map (P), and distance map (Q) into a generator network.

et al. (2022) proposed a SAFRON GAN to address high-resolution histopathology image generation. SAFRON could generate large-size images without requiring same-size inputs. Wei et al. (2019) introduced a GAN to generate synthetic images belonging to diagnostically less common samples to mitigate the data imbalance issue in adenoma detection.

Style Manipulation. Style manipulation and editing has achieved great success in image synthesis and translation tasks (Gong et al., 2021; Huang & Belongie, 2017; Karras et al., 2019; Park et al., 2019). Gong et al. (2021) converted different histopathology images to one target style for minimizing the impact of the various stain distributions. StyleGAN (Karras et al., 2019) proposed an encoder network for encoding style images into the latent vector; and one model could learn multiple image styles from a training set. This network could generate images with various styles while using the pre-trained encoder. Many recent style manipulation GANs such as StyleGAN (Karras et al., 2019), SPADE (Park et al., 2019) applied the style encoding method for effective style transfer.

4.3 SIAN Method

4.3.1 Architecture and Learning Objectives

Figure 4.2 shows the overall architecture of the proposed generator. We employed SPADE (Park et al., 2019) as our baseline architecture, and all the SPADE blocks and SPADE ResNet blocks are replaced with our SIAN blocks and SIAN residual blocks, respectively. Our generator has seven SIAN residual blocks (SIAN ResBlk) and the following up-sampling layer. Each SIAN ResBlk contains two consecutive SIAN blocks followed by ReLU and convolutional layers, the skip connection has a SIAN block, a ReLU, and a convolutional layer. All input factors are down-sampled to the same height and width with the corresponding feature maps in the generator. We follow the same encoder and
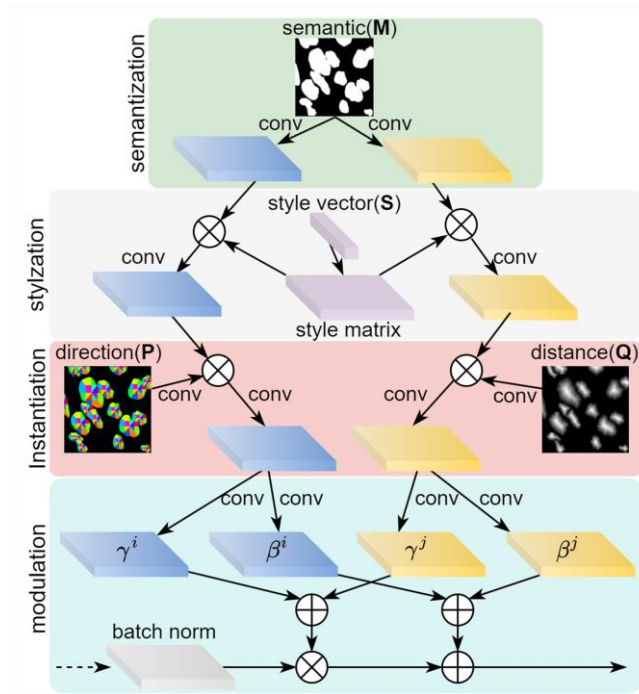
Figure 4.3 SIAN normalization. SIAN block takes four inputs: semantic masks M, style vector S, directional mask P, and distance mask Q for combining different features at multiple phases in the block. $\otimes$ denotes element-wise multiplication, and $\oplus$ is element-wise addition.

discriminator architecture, loss function described in Park et al. (2019). Overall, the loss function contains four components, including the hinge-based conditional adversarial loss (Miyato et al., 2018), the feature matching loss (T.-C. Wang et al., 2018) to minimize the difference between synthetic and real images in the multiscale discriminator, the perceptual loss (Johnson et al., 2016) for minimizing the features between real and synthetic images, the KL divergence loss (Kingma & Welling, 2014) for the encoder to constrain the style vector to the standard Gaussian distribution.

4.3.2 Style-guided Instance-adaptive Normalization

We propose a new conditional normalization block, namely, the Style-guided Instance-Adaptive Normalization (SIAN) to learn instance-level features and integrate image styles for cancers from different organs. Figure 4.3 shows the details of the SIAN block. The block has four phases: semantization, stylization, instantiation, and modulation. The block takes four inputs besides image feature maps, i.e., semantic mask **M**, style vector **S**, direction mask **P**, and distance mask **Q**. The semantization phase embeds image semantics from the input mask; the stylization creates a style matrix from a referred image and integrates image semantics and style. The instantiation phase uses direction and distance maps to distinguish individual nuclei. The modulation phase learns the scale and bias and integrates them into the network. Let **h** denote the input activation of the current layer of the proposed

neural network with a batch size of *N*. Let *H*, *W,* and *C* denote the height, width, and channels of an activation map in *i*th layer. The final modulated activation value ($n \in N$, $c \in C$, $y \in H$, $x \in W$) is defined as

$$\gamma_{c,y,x}(\mathbf{M},\mathbf{S},\mathbf{P},\mathbf{Q})\frac{h_{c,y,x}-\mu_c}{\sigma_c} + \beta_{c,y,x}(\mathbf{M},\mathbf{S},\mathbf{P},\mathbf{Q}) \qquad (4.1)$$

where $h_{c,y,x}$ is the activation output before normalization; the modulation parameter $\gamma_{c,y,x}$ and $\beta_{c,y,x}$ are the elementwise summation of modulation parameters of two branches, i.e., $\gamma_{c,y,x}^{i} + \gamma_{c,y,x}^{j}$ and $\beta_{c,y,x}^{i} + \beta_{c,y,x}^{j}$. μ_c and σ_c are the mean and standard deviation of the activation of channel c, respectively. In the SIAN block, the semantic layout first passes to two convolutional layers, which split the semantic information into two separate branches to learn the directional features and distance features separately. The two branches have the same architecture. In each branch, the convolutional kernel first multiplies with the reshaped style vector, which combines style factors in the block. After that, the instance layouts (direction or distance) are fed through a 1 × 1 convolutional layer and multiplied with the previous convolutional layer. The next convolutional layer learns the compensation of semantic, style, and instance features and then split into two convolutional layers to learn the modulation parameters (*γ* and *β*) spatially. Finally, those modulation parameters and the output of batch normalization are integrated for accurate histopathology image synthesis. All convolutional layers in SIAN use the 3×3 kernel size with 128 filters. Instance masks are applied to generate the semantic mask, and nuclei directional and distance maps. The semantic map is used to separate nuclei and stroma, and the directional and distance maps are useful to demonstrate the boundaries and centroids between two or more touching nuclei. We employed the 2-bin direction mask (Chen et al., 2018) and Medial Axis (MA) distance mask (Wang et al., 2022) as the instance descriptors. The direction map provides important centroid and directional information of nuclei. MA distance mask shows the distance between the nucleus boundary to its skeleton while providing nuclei topological and geometrical features. Figure 4.4 shows two examples. 2-bin direction mask uses two parts of 8 directional regions (8 class insider, 8 class outsider) to obtain directional features, and each class label points to the nucleus centroid. These contain a total number of (16 + 1) classes in the feature map. As shown in Figure 4.4 (c), the inside 8 classes in our direction maps can provide important centroid information of nuclei, which could improve the traditional eight-directional class. MA distance mask shows the distance between the nucleus boundary to its skeleton while providing nuclei topological and geometrical features. It can be observed from Figure 4.4 that the semantic masks lack the ability to show clear boundaries on clustered nuclei, while direction masks and distance masks provide useful important information, e.g., nuclei connectedness, distance, and centroids.
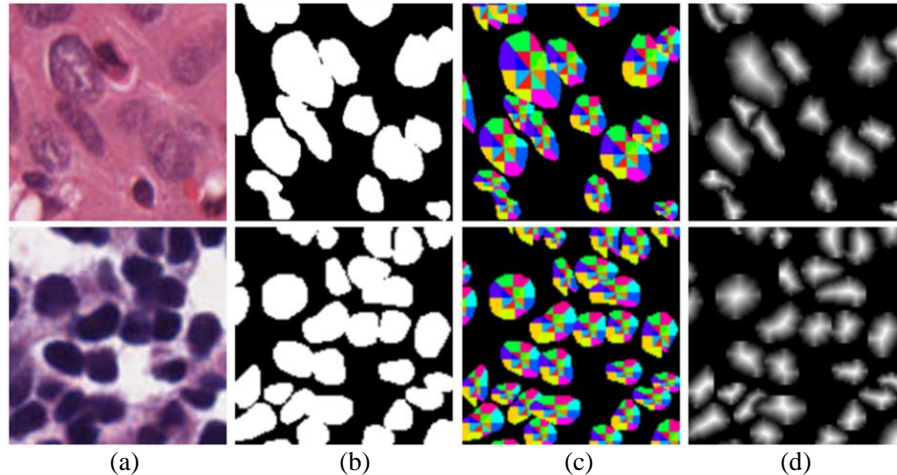
Figure 4.4 Examples of clustered nuclei. a) Real histopathology image patches; b) semantic masks with nuclei in white and stroma in black; c) 2-bin directional masks, which split each nucleus into 16 directional regions (different colors represent different regions); and d) the MA distance masks.

4.4 Experimental Results

4.4.1 Dataset, metrics, and setting

Dataset. The experiments are conducted on the multi-organ nuclei segmentation dataset (MoNuSeg) (Kumar et al., 2019), which has 44 H&E stained histopathology image patches. Both the training and test sets contain images from six organs including breast, liver, kidney, prostate, bladder, and colon; while the training set includes stomach as the seventh organ, and the testing set has brain images.

Evaluation metrics. We employed five metrics to evaluate the method performance for image synthesis, e.g., FID (Heusel et al., 2017), SSIM (Wang et al., 2004), DQ (Kirillov et al., 2019), SQ (Kirillov et al., 2019), and PQ (Kirillov et al., 2019). We used two metrics FID and SSIM to measure the distribution distance and structural similarity between real images and synthetic images, respectively; and used DQ, SQ, and PQ are utilized to assess the nuclei segmentation performance. Specifically, we run a pre-trained segmentation model (SegNet (Badrinarayanan et al., 2017)) which is trained on real images, and then test and evaluate using the synthetic images. In addition, we show the visual comparison of our synthetic images compared to other methods.

Implementation details. The input image size of our approach is 256×256. We used random flip, rotation and median blur for data augmentation. We use the ADAM optimizer with the total training epochs of 50 and batch size of 8 to train the network. The experiments are conducted on an NVIDIA RTX 8000 GPU.

Table 4.1 SIAN: overall performance on MoNuSeg dataset with reconstruction metrics and segmentation metrics

Methods	FID↓	SSIM↑	RQ↑	SQ↑	PQ↑
Pix2pix	170.1	0.467	0.687	0.717	0.493
Sharp-GAN	155.2	0.483	0.721	0.745	0.538
Pix2pixHD	186.3	0.479	0.750	0.753	0.566
SPADE	134.6	0.488	0.705	0.738	0.552
INST	125.4	0.491	0.748	0.768	0.575
STYLE	116.5	0.506	0.743	0.769	0.571
SIAN	**115.7**	**0.515**	**0.757**	**0.761**	**0.586**

During inference, for the style encoder, we take an arbitrary histopathology image as input and output the encoded style vectors. Then, the encoded style vectors together with an arbitrary instance mask are input into the trained generator network to produce histopathology images.

4.4.2 Image Quality Assessment

The proposed method is compared with four state-of-the-art image synthesis models: pix2pixGAN (Isola et al., 2017), Sharp-GAN (Butte et al., 2022), pix2pixHD (T.-C. Wang et al., 2018), and SPADE (Park et al., 2019) using FID, SSIM, DQ, SQ, and PQ metrics. The quantitative results of different approaches on the MoNuSeg test set are shown in Table 4.1. The proposed method outperforms the state-of-the-art methods both in image reconstruction quality using SSIM and FID, and segmentation quality using PQ, SQ, and DQ. In addition, we integrated the instantiation phase (INST) only, and the SIAN block with the style vectors (STYLE), all the evaluation metrics are improved from the baseline SPADE. Overall, we can conclude that our SIAN achieved the best quantitative performance among other methods. Figure 4.5 compares the proposed method, Sharp-GAN, and SPADE using three examples. We noted that Sharp-GAN cannot recover the texture and color distributions of nuclei and stroma in real images, especially in the first and second rows, i.e., the synthetic nuclei have different appearances from real nuclei, and their background stroma lacks meaningful texture and color. SPADE achieved better performance compared to Sharp-GAN, but the generated images were not realistic. Our SIAN generates more realistic images than SPADE and Sharp-GAN. SPADE used the semantic layout as input, while our method used the instance layouts. As shown in Figure 4.6, SPADE tends to generate blur and incorrect nuclei in the clustered region. Our approach produces more accurate boundaries for clustered nuclei.
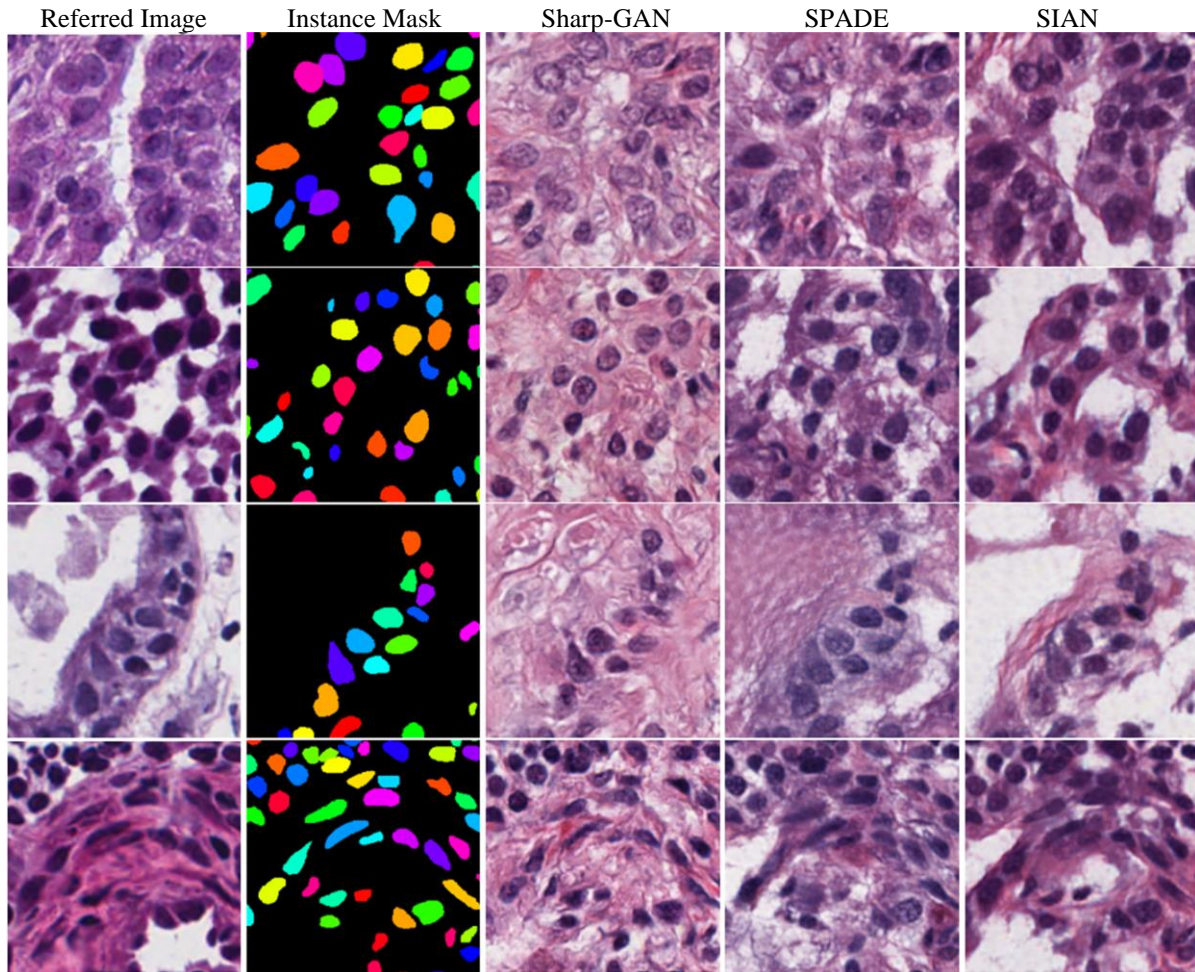
Figure 4.5 Visual comparison of histopathology image synthesis for the MoNuSeg test set.

Table 4.2 SIAN: performance comparison of image synthesis for multiple organs using the FID score.

Methods	Breast	Kidney	Prostate	Bladder	Colon	Lung	Brain
pix2pix	211.0	197.3	237.5	246.9	239.3	**187.5**	304.4
Sharp-GAN	**198.0**	190.4	235.3	211.4	208.6	195.4	260.3
pix2pixHD	260.0	222.1	260.1	278.6	221.8	252.8	278.6
SPADE	212.4	187.9	200.4	212.5	222.3	207.2	245.5
SIAN	200.4	**173.9**	**190.2**	**196.9**	**204.1**	194.6	**239.6**
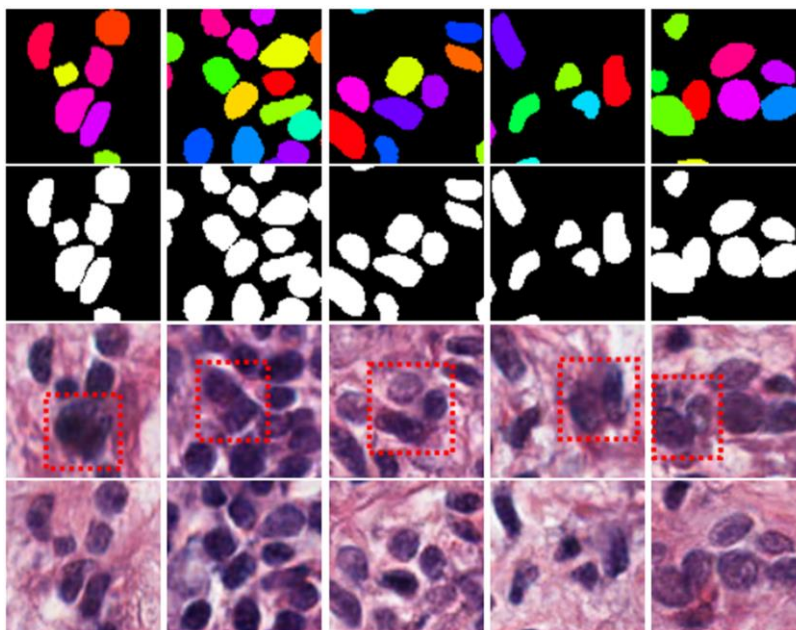
Figure 4.6 Synthesis for clustered nuclei. First row: instance masks (different colors represent different nuclei). Second row: semantic masks. Third row: results of SPADE. Fourth row: results of SIAN.

4.4.3 Multi-organ Image Synthesis

To evaluate synthetic images across multiple organs, we compare the generation performance of four methods for each organ using FID scores. The results are shown in Table 4.2. The proposed method outperforms other state-of-the-art methods in kidney, prostate, bladder, colon, and brain images, and achieved the second-best results for synthesizing breast and lung images. Figure 4.1 shows the results of SIAN across multiple organs, the color and texture distribution of foreground nuclei and background stroma is close to the reals.

4.4.4 Nuclei Segmentation using Synthetic Images

In this experiment, we evaluate the effectiveness of synthetic augmentation for training segmentation networks. We train SegNet (Badrinarayanan et al. (2017)) with different input configurations (as shown in Table 4.3). In experiments, nucleus-like polygons are generated as the synthetic nuclei instance masks (Hou et al., 2019), in total 5,000 synthetic instance masks are generated and applied to produce corresponding semantic, directional, and distance masks. Then, the pre-trained SIAN is used to apply seven different style vectors encoded from seven different organs (around 700 synthetic images per organ) and generate realistic histopathology images. Finally, we test and evaluate the segmentation performance with the MoNuSeg test set using DQ, SQ, and PQ metrics. We compared the proposed approach to other methods on synthetic augmentation. Synthetic images generated from other methods follow their design. As shown in Table 4.3, with synthetic training images from pix2pixHD, Sharp-

Table 4.3 SIAN: performance of SegNet using different training sets. ∗ denotes the training set augmented using traditional augmentation techniques, e.g., flip, rotate, blur. '+method' denotes applying synthetic augmentation∗ with 5,000 synthetic images generated from 'method' to the training set.

Training Set	RQ	SQ	PQ
MoNuSeg training set	0.704	0.737	0.521
MoNuSeg training set*	0.732	0.739	0.538
+pix2pixHD	0.742	0.737	0.544
+Sharp-GAN	0.740	0.739	0.547
+SPADE	0.743	0.738	0.549
+SIAN	**0.748**	**0.742**	**0.555**

GAN, and SPADE, the performance of nuclei segmentation could be significantly improved. The proposed SIAN help generate the best segmentation performance.

4.5 SIAN Conclusion

In this study, we propose style-guided instance-adaptive normalization (SIAN) for multi-organ histopathology image synthesis, which integrates instance layouts and encodes style vectors into a generative network. SIAN synthesizes histopathology images with styles that align with the image styles of multiple organs. SIAN utilizes the directional and distance masks from the nuclei instance maps and generates clear boundaries for densely-clustered nuclei. With the integration of the stylization phase, SIAN allows style editing for synthesizing images of multiple organs. In addition, SIAN demonstrates its effectiveness in augmenting the training set and improving the overall performance of a deep learning model for nuclei segmentation.

Chapter 5: Application to other imaging modalities and problems

Wang, H., Xu, F., Cai, L., Salvato, D., Capriotti, L., Yao, T., & Xian, M. (2023) A Fine Pore-preserved Deep Neural Network for Porosity Analytics of a High Burnup U-10Zr Metallic Fuel.

5.1 Application to Other Image Modalities

Multi-task learning (MTL) has shown great promise in various image modalities and image processing applications, including satellite images (Bischke et al., 2019; Liu et al., 2018; Pandey et al., 2018), ultrasound images (Chowdary et al., 2022; Shareef et al., 2023; Zhou et al., 2021), computerized tomography (CT) images (He et al., 2020; Zhai et al., 2020), natural images (Abdulnabi et al., 2015; Huang et al., 2021; Wang et al., 2009; Yim et al., 2015; Zhang et al., 2014; Zhao et al., 2018), and scene images (Ishihara et al., 2021; Jha et al., 2020; Kendall et al., 2018; Zhang et al., 2021). For example, Abdulnabi et al. (2015) proposed a multi-task convolutional neural network (CNN) that predicts each data attribute using multiple attribute-specific subnets for accurate image recognition. This model can simultaneously share visual knowledge among different tasks; under-sampled attributes can leverage shared knowledge from other classifiers to improve their performance. Zhang et al. (2021) proposed a DPSNet that performed depth estimation, camera pose estimation and semantic scene segmentation simultaneously for accurate scenario understanding. Shareef et al. (2023) presented a multi-task learning approach that utilized tumor segmentation as a regularizer and an auxiliary task to improve the representation and performance of the primary task of tumor classification for breast ultrasound imaging. Bischke et al. (2019) introduced a multi-task learning network and a loss that predicts the distance map and semantic map to preserve the accurate building footprints in satellite images. Numerous real-world applications increasingly use MTL to boost their performance and improve generalizability by training a uniform model to predict multiple tasks simultaneously. Due to the space limit, more MTL applications can refer to the survey paper, such as Crawshaw (2020), Ruder (2017), Vandenhende et al. (2021).

It is worth noting that in this study, we focus on deep learning approaches for MTL in computer vision applications. For a comprehensive overview of MTL in other application domains, the reader could refer to the relevant literature, such as natural language processing (A. Wang et al., 2018; Worsham & Kalita, 2020), speech recognition (Deng et al., 2013), biomedical text mining (Peng et al., 2020), bioinformatics (Widmer et al., 2010), chemoinformatics (Sosnin et al., 2019), web search (Chapelle et al., 2010), reinforcement learning (Vithayathil Varghese & Mahmoud, 2020), etc.
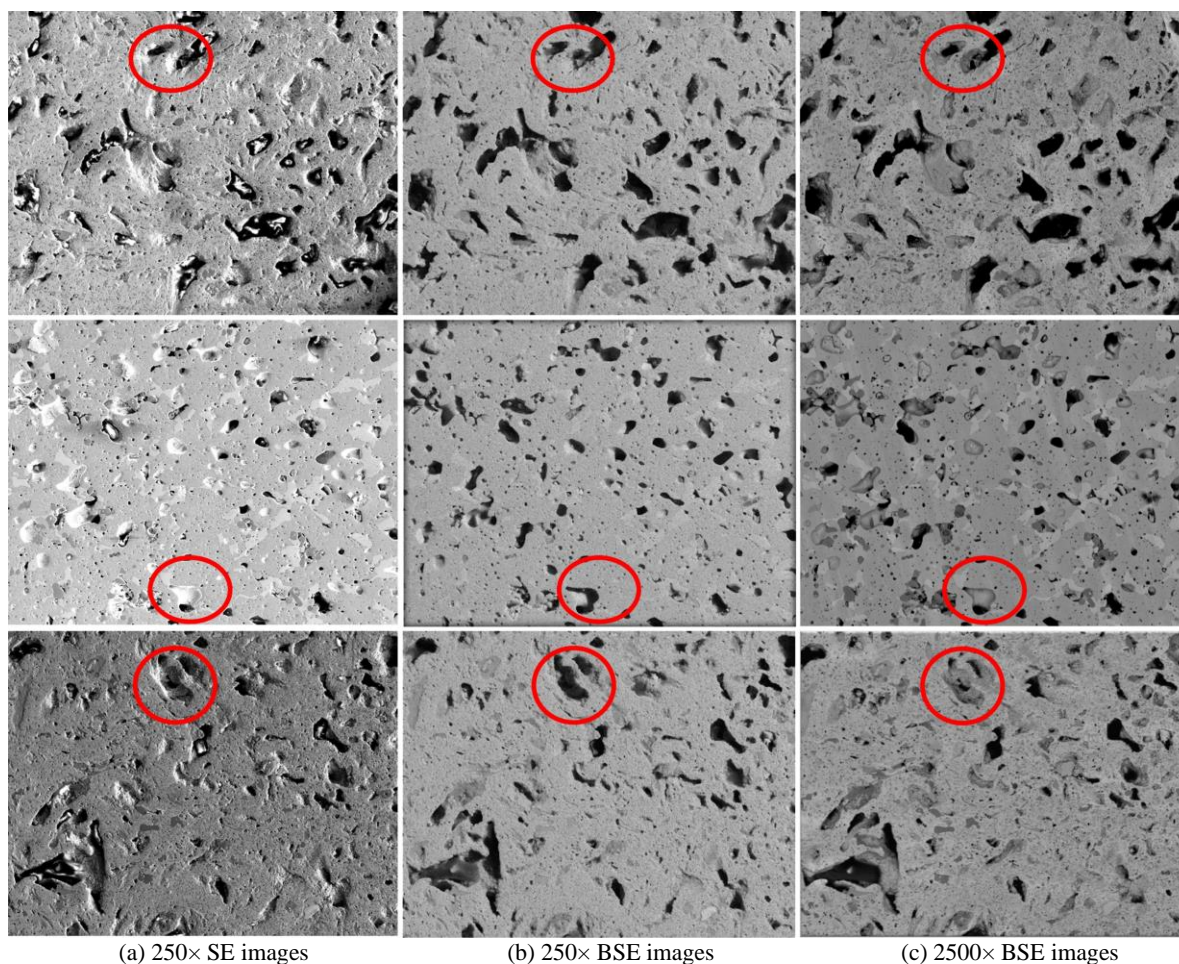
(a) 250× SE images   (b) 250× BSE images   (c) 2500× BSE images

Figure 5.1 Images patches of U-10Zr fuel from Secondary Electron Microscop (SE) and Backscattered Electron (BSE) images captured at different magnifications. In each row, red circles highlight the target microstructure in different image formats and magnification, the same microstructures could have large variance in color and textures.

5.2 Application to Scanning Electron Microscopy (SEM) Imaging

Advanced scanning electron microscopy SEM image analysis technologies have enabled researchers to gain a quantitative understanding of the irradiation behavior of metallic fuel, including Zr redistribution, FCMI, and FCCI (Benson et al., 2021; Carmack et al., 2016; Harp et al., 2018; Harp et al., 2017; Liu et al., 2021; Salvato et al., 2022; Yao et al., 2020). The traditional approaches to learning metallic fuel in SEM imaging using image processing, such as image thresholding, to manually select and measure the microstructure of interest, are prone to human bias and inconsistent operation during the inference. Furthermore, the presence of multiple sources, scales, and diverse image conditions can introduce significant morphological variation across different microstructures, and color and texture variance. It is challenging to accurately detect and segment the microstructures in SEM images, such as pores and Zr. Figure 5.1 highlights how the same microstructure can exhibit large color and texture

(a) Image   (b) Ground truth   (c) SegNet   (d) U-Net   (e) Ours
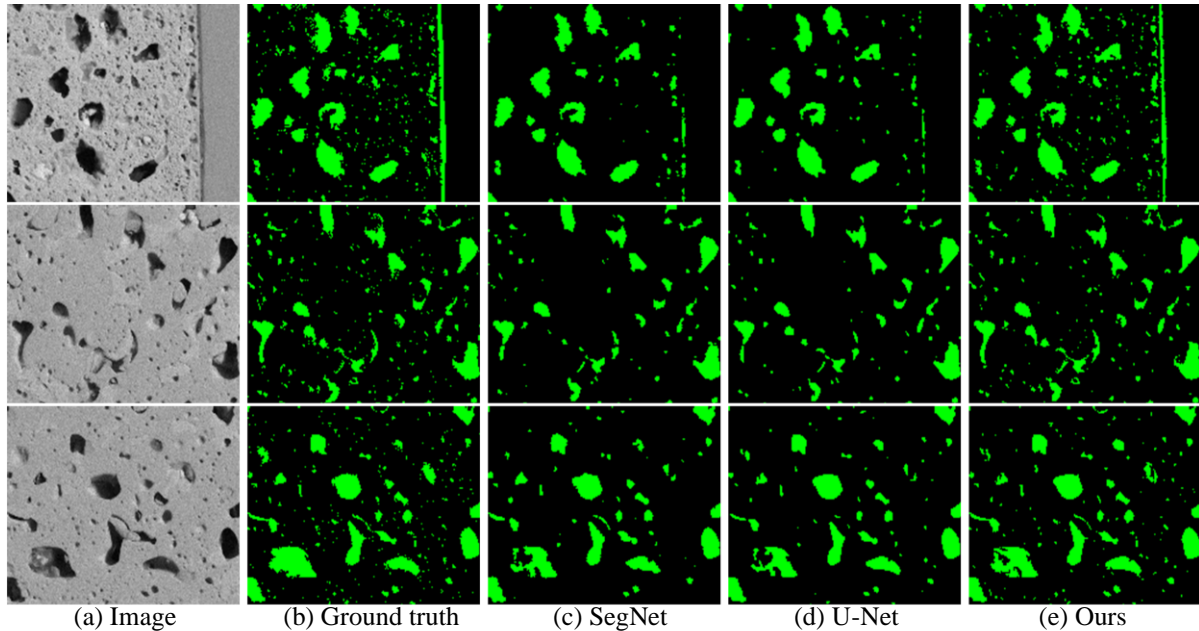
Figure 5.2 The visual representation of pore segmentation comparison with state-of-the-art methods on BSE images.

variance at different image types and magnifications. Thus, an automatic, robust, and generalizable model for accurately detecting microstructures in SEM images is highly desired in SEM image analysis.

In this research, we proposed a novel deep-learning architecture for analyzing porosity in SEM examination. The proposed network utilized the Bend-Net architecture (Chapter 2) as a backbone network for segmenting pores from SEM images. Due to the majority of neural network architecture following the previous Bend-Net work and quantitative analysis of porosity for metallic fuel is out of the histopathological domain. The detailed description of the method and the quantitative analysis can be found in Wang, Xu, et al. (2023). From the overall qualitative results, Figure 5.2 shows the comparison of our method to two benchmark semantic segmentation methods, SegNet (Badrinarayanan et al., 2017), U-Net (Ronneberger et al., 2015). It can observe that our results are closer to the ground truth, and the prediction of the SegNet and U-Net have high false negatives compared to our methods (low recall), especially for small pores detection and the pores with irregular shapes.

Chapter 6: Conclusion

Through the thesis, we discuss the applications of multi-task learning for addressing the challenges in histopathology image processing. We explore the solutions of deep neural networks and multi-task learning for histopathology nuclei segmentation, and histopathology gland segmentation. We propose novel multi-task learning networks with task-specific objective functions to model tissue topological differences. We introduce the generative model for synthesizing realistic histopathology images and enabling histopathological style transfer in images. Meanwhile, our research in histopathology image analysis can be adapted and extended into other image domains and modalities. The proposed morphological and topological knowledge on the tissue patterns can be applied to other image patterns, such as city roads and human movements. The research experience that I have gained through the Ph. D. studies includes analyzing the data, identifying the issues, collecting useful information, creating plans and steps to solve the problems, sharing insights, and providing the conclusion and recommendations.

6.1 List of Publications

Wang, H., Xian, M., Vakanski, A., & Shareef, B. (2023). SIAN: Style-Guided Instance-Adaptive Normalization for Multi-Organ Histopathology Image Synthesis. IEEE International Symposium on Biomedical Imaging (ISBI),

Wang, H., Xian, M., & Vakanski, A. (2022). TA-net: Topology-aware network for gland segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,

Wang, H., Vakanski, A., Shi, C., & Xian, M. (2021). Bend-Net: Bending Loss Regularized Multi-task Learning Network for Nuclei Segmentation in Histopathology Images. *arXiv preprint arXiv:2109.15283*.

Wang, H., Xian, M., & Vakanski, A. (2020). Bending loss regularized network for nuclei segmentation in histopathology images. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI),

Wang, H., Xu, F., Cai, L., Salvato, D., Capriotti, L., Yao, T., & Xian, M. (2023) A Fine Pore-preserved Deep Neural Network for Porosity Analytics of a High Burnup U-10Zr Metallic Fuel. (under review)

Wang, H, Xian, M., Vakanski, A., FATANet: Feedback Attention Topology-aware Network with Glandular Structures for Histopathology Gland Segmentation. (under review)

Butte, S.*, **Wang, H.***, Vakanski, A., & Xian, M. (2023). Enhanced Sharp-GAN For Histopathology Image Synthesis. IEEE International Symposium on Biomedical Imaging (ISBI) (*Co-first author contributed equally),

Butte, S.*, **Wang, H.***, Xian, M., & Vakanski, A. (2022). Sharp-GAN: Sharpness Loss Regularized GAN for Histopathology Image Synthesis. IEEE International Symposium on Biomedical Imaging (ISBI) (*Co-first author contributed equally),

Shi, C., Xian, M., Zhou, X., **Wang, H.**, & Heng-DaCheng. (2021). Multi-Slice Low-Rank Tensor Decomposition Based Multi-Atlas Segmentation: Application to Automatic Pathological Liver CT Segmentation. *Medical Image Analysis*, *73*, 102152.

Butte, S., Vakanski, A., Duellman, K., **Wang, H.**, & Mirkouei, A. (2021). Potato Crop Stress Identification in Aerial Images using Deep Learning-based Object Detection. *Agronomy Journal*, *41*(1).

Bibliography

Abdulnabi, A. H., Wang, G., Lu, J., & Jia, K. (2015). Multi-task CNN model for attribute prediction. *IEEE Transactions on Multimedia*, *17*(11), 1949-1959.

Alemi Koohbanani, N., Jahanifar, M., Gooya, A., & Rajpoot, N. (2019). Nuclear instance segmentation using a proposal-free spatially aware deep learning framework. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22,

Ali, S., & Madabhushi, A. (2012). An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE transactions on medical imaging*, *31*(7), 1448-1460.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481-2495.

Bai, M., & Urtasun, R. (2017). Deep watershed transform for instance segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition,

Benson, M. T., Harp, J. M., Xie, Y., Yao, T., Tolman, K. R., Wright, K. E., King, J. A., Hawari, A. I., & Cai, Q. (2021). Out-of-pile and postirradiated examination of lanthanide and lanthanide-palladium interactions for metallic fuel. *Journal of Nuclear Materials*, *544*, 152727.

Bergou, M., Wardetzky, M., Robinson, S., Audoly, B., & Grinspun, E. (2008). Discrete elastic rods. In *ACM SIGGRAPH 2008 papers* (pp. 1-12).

Bischke, B., Helber, P., Folz, J., Borth, D., & Dengel, A. (2019). Multi-task learning for segmentation of building footprints with deep neural networks. 2019 IEEE International Conference on Image Processing (ICIP),

Bloom, H. J. G., & Richardson, W. (1957). Histological grading and prognosis in breast cancer: a study of 1409 cases of which 359 have been followed for 15 years. *British journal of cancer*, *11*(3), 359.

Blum, H. (1967). A transformation for extracting new descriptions of shape. *Models for the perception of speech and visual form*, 362-380.

Butte, S., Vakanski, A., Duellman, K., Wang, H., & Mirkouei, A. (2021). Potato Crop Stress Identification in Aerial Images using Deep Learning-based Object Detection. *Agronomy Journal*, *41*(1).

Butte, S., Wang, H., Vakanski, A., & Xian, M. (2023). Enhanced Sharp-GAN For Histopathology Image Synthesis. IEEE International Symposium on Biomedical Imaging (ISBI),

Butte, S., Wang, H., Xian, M., & Vakanski, A. (2022). Sharp-gan: Sharpness loss regularized gan for histopathology image synthesis. 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI),

Canham, P. B. (1970). The minimum energy of bending as a possible explanation of the biconcave shape of the human red blood cell. *Journal of theoretical biology*, *26*(1), 61-81.

Carmack, W. J., Chichester, H. M., Porter, D. L., & Wootan, D. W. (2016). Metallography and fuel cladding chemical interaction in fast flux test facility irradiated metallic U-10Zr MFF-3 and MFF-5 fuel pins. *Journal of Nuclear Materials*, *473*, 167-177.

Caruana, R. (1998). *Multitask learning*. Springer.

Chapelle, O., Shivaswamy, P., Vadrevu, S., Weinberger, K., Zhang, Y., & Tseng, B. (2010). Multi-task learning for boosting with application to web search ranking. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining,

Chen, H., Qi, X., Yu, L., & Heng, P.-A. (2016). DCAN: deep contour-aware networks for accurate gland segmentation. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition,

Chen, L.-C., Hermans, A., Papandreou, G., Schroff, F., Wang, P., & Adam, H. (2018). Masklab: Instance segmentation by refining object detection with semantic and direction features. Proceedings of the IEEE conference on computer vision and pattern recognition,

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834-848.

Cheng, J., & Rajapakse, J. C. (2008). Segmentation of clustered nuclei with shape markers and marking function. *IEEE Transactions on Biomedical Engineering*, *56*(3), 741-748.

Chowdary, J., Yogarajah, P., Chaurasia, P., & Guruviah, V. (2022). A multi-task learning framework for automated segmentation and classification of breast tumors from ultrasound images. *Ultrasonic imaging*, *44*(1), 3-12.

Clough, J. R., Byrne, N., Oksuz, I., Zimmer, V. A., Schnabel, J. A., & King, A. P. (2020). A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(12), 8766-8778.

Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009.09796*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition,

Deng, L., Hinton, G., & Kingsbury, B. (2013). New types of deep neural network learning for speech recognition and related applications: An overview. 2013 IEEE international conference on acoustics, speech and signal processing,

Deshpande, S., Minhas, F., Graham, S., & Rajpoot, N. (2022). SAFRON: Stitching across the frontier network for generating colorectal cancer histology images. *Medical Image Analysis*, *77*, 102337.

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

Ding, H., Pan, Z., Cen, Q., Li, Y., & Chen, S. (2020). Multi-scale fully convolutional network for gland segmentation using three-class classification. *Neurocomputing*, *380*, 150-161.

Du, Q., Liu, C., & Wang, X. (2006). Simulating the deformation of vesicle membranes under elastic bending energy in three dimensions. *Journal of computational physics*, *212*(2), 757-777.

Duncan, J. S., Lee, F. A., Smeulders, A. W., & Zaret, B. L. (1991). A bending energy model for measurement of cardiac shape deformity. *IEEE transactions on medical imaging*, *10*(3), 307-320.

Epstein, J. I., Allsbrook Jr, W. C., Amin, M. B., Egevad, L. L., & Committee, I. G. (2005). The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *The American journal of surgical pathology*, *29*(9), 1228-1242.

Estrada, R., Tomasi, C., Schmidler, S. C., & Farsiu, S. (2014). Tree topology estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(8), 1688-1701.

Fleming, M., Ravula, S., Tatishchev, S. F., & Wang, H. L. (2012). Colorectal carcinoma: Pathologic aspects. *Journal of gastrointestinal oncology*, *3*(3), 153.

Gao Huang, Z. L., Laurens Van Der Maaten, and Kilian Q Weinberger. (2017). Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Gong, X., Chen, S., Zhang, B., & Doermann, D. (2021). Style consistent image generation for nuclei instance segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,

Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.-A., Snead, D., Tsang, Y. W., & Rajpoot, N. (2019). MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis*, *52*, 199-211.

Graham, S., & Rajpoot, N. M. (2018). SAMS-NET: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018),

Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2019). Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, *58*, 101563.

Harp, J. M., Capriotti, L., Chichester, H. J., Medvedev, P. G., Porter, D. L., & Hayes, S. L. (2018). Postirradiation examination on metallic fuel in the AFC-2 irradiation test series. *Journal of Nuclear Materials*, *509*, 454-464.

Harp, J. M., Porter, D. L., Miller, B. D., Trowbridge, T. L., & Carmack, W. J. (2017). Scanning electron microscopy examination of a Fast Flux Test Facility irradiated U-10Zr fuel cross section clad with HT-9. *Journal of Nuclear Materials*, *494*, 227-239.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,

He, T., Hu, J., Song, Y., Guo, J., & Yi, Z. (2020). Multi-task learning for the segmentation of organs at risk with label dependence. *Medical Image Analysis*, *61*, 101666.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, *30*.

Hou, L., Agarwal, A., Samaras, D., Kurc, T. M., Gupta, R. R., & Saltz, J. H. (2019). Robust histopathology image analysis: To label or to synthesize? Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Hu, X., Li, F., Samaras, D., & Chen, C. (2019). Topology-preserving deep image segmentation. *Advances in neural information processing systems*, *32*.

Huang, X., & Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of the IEEE international conference on computer vision,

Huang, Z., Zhang, J., & Shan, H. (2021). When age-invariant face recognition meets face age synthesis: A multi-task learning framework. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Ishihara, K., Kanervisto, A., Miura, J., & Hautamaki, V. (2021). Multi-task learning with attention for end-to-end autonomous driving. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition,

Jha, A., Kumar, A., Banerjee, B., & Chaudhuri, S. (2020). Adamt-net: An adaptive weight learning based multi-task learning model for scene understanding. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,

Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14,

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. Proceedings of the IEEE conference on computer vision and pattern recognition,

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. International Conference on Learning Representations,

Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O. F., Tsougenis, E., Chen, H., Heng, P.-A., Li, J., & Hu, Z. (2019). A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, *39*(5), 1380-1391.

Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., & Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, *36*(7), 1550-1560.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13,

Liu, H., Fu, Z., Han, J., Shao, L., & Liu, H. (2018). Single satellite imagery simultaneous super-resolution and colorization using multi-task deep neural networks. *Journal of Visual Communication and Image Representation*, *53*, 20-30.

Liu, X., Capriotti, L., Yao, T., Harp, J. M., Benson, M. T., Wang, Y., Teng, F., & He, L. (2021). Fuel-cladding chemical interaction of a prototype annular U-10Zr fuel with Fe-12Cr ferritic/martensitic HT-9 cladding. *Journal of Nuclear Materials*, *544*, 152588.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition,

Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*.

Montironi, R., Mazzuccheli, R., Scarpelli, M., Lopez‐Beltran, A., Fellegara, G., & Algaba, F. (2005). Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies. *BJU international*, *95*(8), 1146-1152.

Mosinska, A., Marquez-Neila, P., Koziński, M., & Fua, P. (2018). Beyond the pixel-wise loss for topology-aware delineation. Proceedings of the IEEE conference on computer vision and pattern recognition,

Naylor, P., Laé, M., Reyal, F., & Walter, T. (2018). Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, *38*(2), 448-459.

Nguyen, K., Jain, A. K., & Allen, R. L. (2010). Automated gland segmentation and classification for gleason grading of prostate tissue images. 2010 20th International Conference on Pattern Recognition,

Noble, J. H., & Dawant, B. M. (2011). An atlas-navigated optimal medial axis and deformable model algorithm (nomad). *Medical image analysis*, *15*(6), 877-884.

Oda, H., Roth, H. R., Chiba, K., Sokolić, J., Kitasaka, T., Oda, M., Hinoki, A., Uchida, H., Schnabel, J. A., & Mori, K. (2018). BESNet: boundary-enhanced segmentation of cells in histopathological images. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11,

Pandey, S., Agarwal, T., & Krishnan, N. C. (2018). Multi-task deep learning for predicting poverty from satellite images. Proceedings of the AAAI Conference on Artificial Intelligence,

Park, T., Liu, M.-Y., Wang, T.-C., & Zhu, J.-Y. (2019). Semantic image synthesis with spatially-adaptive normalization. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition,

Paul, A., & Mukherjee, D. P. (2016). Gland segmentation from histology images using informative morphological scale space. 2016 IEEE International Conference on Image Processing (ICIP),

Peng, Y., Chen, Q., & Lu, Z. (2020). An empirical study of multi-task learning on BERT for biomedical text mining. *arXiv preprint arXiv:2005.02799*.

Qu, H., Yan, Z., Riedlinger, G. M., De, S., & Metaxas, D. N. (2019). Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22,

Raza, S. E. A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., Khan, M., & Rajpoot, N. M. (2019). Micro-Net: A unified model for segmentation of various objects in microscopy images. *Medical Image Analysis*, *52*, 160-173.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18,

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Salvato, D., Liu, X., Murray, D. J., Paaren, K. M., Xu, F., Pavlov, T., Benson, M. T., Capriotti, L., & Yao, T. (2022). Transmission electron microscopy study of a high burnup U-10Zr metallic fuel. *Journal of Nuclear Materials*, *570*, 153963.

Sebastian, T. B., Klein, P. N., & Kimia, B. B. (2001). Recognition of shapes by editing shock graphs. ICCV,

Shaban, M., Awan, R., Fraz, M. M., Azam, A., Tsang, Y.-W., Snead, D., & Rajpoot, N. M. (2020). Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE transactions on medical imaging*, *39*(7), 2395-2405.

Shareef, B. M., Xian, M., Sun, S., Vakanski, A., Ding, J., Ning, C., & Cheng, H.-D. (2023). A Benchmark for Breast Ultrasound Image Classification. *Available at SSRN 4339660*.

Shi, C., Xian, M., Zhou, X., Wang, H., & Heng-DaCheng. (2021). Multi-Slice Low-Rank Tensor Decomposition Based Multi-Atlas Segmentation: Application to Automatic Pathological Liver CT Segmentation. *Medical Image Analysis*, *73*, 102152.

Shit, S., Paetzold, J. C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J. P., Bauer, U., & Menze, B. H. (2021). clDice-a novel topology-preserving loss function for tubular structure segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Shotton, J., Girshick, R., Fitzgibbon, A., Sharp, T., Cook, M., Finocchio, M., Moore, R., Kohli, P., Criminisi, A., & Kipman, A. (2012). Efficient human pose estimation from single depth images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(12), 2821-2840.

Siddiqi, K., Shokoufandeh, A., Dickenson, S., & Zucker, S. W. (1998). Shock graphs and shape matching. Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271),

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2022). Cancer statistics, 2022. *CA: a cancer journal for clinicians*, *72*(1), 7-33.

Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P.-A., Guo, Y. B., Wang, L. Y., Matuszewski, B. J., Bruni, E., & Sanchez, U. (2017). Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis*, *35*, 489-502.

Sirinukunwattana, K., Snead, D. R., & Rajpoot, N. M. (2015). A stochastic polygons model for glandular structures in colon histology images. *IEEE transactions on medical imaging*, *34*(11), 2366-2378.

Sosnin, S., Vashurina, M., Withnall, M., Karpov, P., Fedorov, M., & Tetko, I. V. (2019). A survey of multi‐task learning methods in chemoinformatics. *Molecular informatics*, *38*(4), 1800108.

Stuhmer, J., Schroder, P., & Cremers, D. (2013). Tree shape priors with connectivity constraints using convex relaxation on general graphs. Proceedings of the IEEE International conference on Computer Vision,

Tosun, A. B., & Gunduz-Demir, C. (2010). Graph run-length matrices for histopathological image segmentation. *IEEE transactions on medical imaging*, *30*(3), 721-732.

Tsogkas, S., & Dickinson, S. (2017). AMAT: Medial axis transform for natural images. Proceedings of the IEEE International Conference on Computer Vision,

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A. M., Esposito, I., & Navab, N. (2016). Structure-preserving color normalization and sparse stain separation for histological images. *IEEE transactions on medical imaging*, *35*(8), 1962-1971.

Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D., & Van Gool, L. (2021). Multi-task learning for dense prediction tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3614-3633.

Verbeek, P., & Van Vliet, L. (1993). Curvature and bending energy in digitized 2D and 3D images. 8th Scandinavian Conference on Image Analysis, Tromso, Norway,

Veta, M., Pluim, J. P., Van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, *61*(5), 1400-1411.

Vithayathil Varghese, N., & Mahmoud, Q. H. (2020). A survey of multi-task deep reinforcement learning. *Electronics*, *9*(9), 1363.

Vu, Q. D., Graham, S., Kurc, T., To, M. N. N., Shaban, M., Qaiser, T., Koohbanani, N. A., Khurram, S. A., Kalpathy-Cramer, J., & Zhao, T. (2019). Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology*, 53.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*.

Wang, H., Vakanski, A., Shi, C., & Xian, M. (2021). Bend-Net: Bending Loss Regularized Multitask Learning Network for Nuclei Segmentation in Histopathology Images. *arXiv preprint arXiv:2109.15283*.

Wang, H., Xian, M., & Vakanski, A. (2020). Bending loss regularized network for nuclei segmentation in histopathology images. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI),

Wang, H., Xian, M., & Vakanski, A. (2022). Ta-net: Topology-aware network for gland segmentation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision,

Wang, H., Xian, M., Vakanski, A., & Shareef, B. (2023). SIAN: Style-Guided Instance-Adaptive Normalization for Multi-Organ Histopathology Image Synthesis. IEEE International Symposium on Biomedical Imaging (ISBI),

Wang, H., Xu, F., Cai, L., Salvatob, D., Capriottib, L., Yao, T., & Xian, M. (2023). A Fine Pore-preserved Deep Neural Network for Porosity Analytics of a High Burnup U-10Zr Metallic Fuel. *arXiv preprint arXiv:*.

Wang, P., Li, Y., Singh, K. K., Lu, J., & Vasconcelos, N. (2021). Imagine: Image synthesis by image-guided model inversion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. Proceedings of the IEEE conference on computer vision and pattern recognition,

Wang, X., Zhang, C., & Zhang, Z. (2009). Boosted multi-task learning for face verification with applications to web image and video search. 2009 IEEE conference on computer vision and pattern recognition,

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600-612.

Wei, J., Suriawinata, A., Vaickus, L., Ren, B., Liu, X., Wei, J., & Hassanpour, S. (2019). Generative image translation for data augmentation in colorectal histopathology images. *Proceedings of machine learning research*, *116*, 10.

Widmer, C., Leiva, J., Altun, Y., & Rätsch, G. (2010). Leveraging sequence classification by taxonomy-based multitask learning. Research in Computational Molecular Biology: 14th Annual International Conference, RECOMB 2010, Lisbon, Portugal, April 25-28, 2010. Proceedings 14,

Worsham, J., & Kalita, J. (2020). Multi-task learning for natural language processing in the 2020s: where are we going? *Pattern Recognition Letters*, *136*, 120-126.

Xie, Y., Zhang, J., Liao, Z., Verjans, J., Shen, C., & Xia, Y. (2021). Intra-and inter-pair consistency for semi-supervised gland segmentation. *IEEE Transactions on Image Processing*, *31*, 894-905.

Xing, F., Xie, Y., & Yang, L. (2015). An automatic learning-based framework for robust nucleus segmentation. *IEEE transactions on medical imaging*, *35*(2), 550-566.

Xu, Y., Li, Y., Wang, Y., Liu, M., Fan, Y., Lai, M., Eric, I., & Chang, C. (2017). Gland instance segmentation using deep multichannel neural networks. *IEEE Transactions on Biomedical Engineering*, *64*(12), 2901-2912.

Xue, Y., Ye, J., Zhou, Q., Long, L. R., Antani, S., Xue, Z., Cornwell, C., Zaino, R., Cheng, K. C., & Huang, X. (2021). Selective synthetic augmentation with HistoGAN for improved histopathology image classification. *Medical Image Analysis*, *67*, 101816.

Yan, Z., Yang, X., & Cheng, K.-T. (2020). Enabling a single deep learning model for accurate gland instance segmentation: A shape-aware adversarial learning framework. *IEEE transactions on medical imaging*, *39*(6), 2176-2189.

Yang, X., Li, H., & Zhou, X. (2006). Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *53*(11), 2405-2414.

Yao, T., Capriotti, L., Harp, J. M., Liu, X., Wang, Y., Teng, F., Murray, D. J., Winston, A. J., Gan, J., & Benson, M. T. (2020). α-U and ω-UZr2 in neutron irradiated U-10Zr annular metallic fuel. *Journal of Nuclear Materials*, *542*, 152536.

Yim, J., Jung, H., Yoo, B., Choi, C., Park, D., & Kim, J. (2015). Rotating your face using multi-task deep neural network. Proceedings of the IEEE conference on computer vision and pattern recognition,

Young, I. T., Walker, J. E., & Bowie, J. E. (1974). An analysis technique for biological shape. I. *Information and control*, *25*(4), 357-370.

Zeng, Z., Xie, W., Zhang, Y., & Lu, Y. (2019). RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images. *Ieee Access*, *7*, 21420-21428.

Zhai, P., Tao, Y., Chen, H., Cai, T., & Li, J. (2020). Multi-task learning for lung nodule classification on chest CT. *Ieee Access*, *8*, 180317-180327.

Zhang, J., Su, Q., Tang, B., Wang, C., & Li, Y. (2021). Dpsnet: Multitask learning using geometry reasoning for scene depth and semantics. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014). Facial landmark detection by deep multi-task learning. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13,

Zhao, X., Li, H., Shen, X., Liang, X., & Wu, Y. (2018). A modulation module for multi-task learning with applications in image retrieval. Proceedings of the European Conference on Computer Vision (ECCV),

Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.-T., & Shen, D. (2021). Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Medical Image Analysis*, *70*, 101918.

Zhou, Y., Onder, O. F., Dou, Q., Tsougenis, E., Chen, H., & Heng, P.-A. (2019). Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26,